



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Laure Guitton-Sert

le vendredi 11 décembre 2015

Titre :

Identification de nouveaux mécanismes de régulation temporelle des origines
de réplication dans les cellules humaines

École doctorale et discipline ou spécialité :

ED BSB : Cancérologie

Unité de recherche :

CRCT - Centre de Recherche en Cancérologie de Toulouse INSERM UMR1037

Directeur/trice(s) de Thèse :

Jean-Sébastien HOFFMANN

Jury :

Pr. Michelle DEBATISSE - Rapportrice

Dr. Bernard LOPEZ - Rapporteur

Dr. Gaëlle LEGUBE - Examinatrice

Dr. Philippe COULOMBE - Examineur

Pr. Kerstin BYSTRIKY - Présidente du jury

Anne VIDAL-FERNANDEZ - Invitée

A ma grand-mère, à Martine, à Brigitte, à Tonton Yvan, à Bob, à François...

*A mon **Super** Papi, parti car « c'est la vie ! », laissant un vide vertigineux derrière lui.*

REMERCIEMENTS

Je voudrais commencer par remercier évidemment Anne, qui m'a encadrée pendant 3 belles années. Tu as fait un tel discours à la fin de ma soutenance que j'en ai été toute bouleversée et n'ai pu en retour faire mes remerciements. Je tiens donc à te remercier tout particulièrement pour tout ce que tu m'as enseigné, à la paillasse en termes de manips, d'organisation et de rigueur, au bureau en termes de biblio, de présentation des résultats, à la machine café, en terme de maternité et de jeune scientifique! Il y a quand même un truc que tu n'auras pas réussi à m'inculquer tellement j'y mets de la mauvaise volonté, c'est le cahier de labo... Pourtant, on aura essayé. Je n'oublierai jamais non plus que tu m'as emmenée à Cold Spring Harbor Laboratory, ni cette petite virée dans les rues de New York, et en particulier ces délicieuses pancakes au bord de Central Park ! J'ai beaucoup regretté que tu aies eu à partir pour d'absurdes raisons, et je suis plus qu'heureuse que tu aies eu ce poste de maître de conf que tu mérites tant. Quand je l'ai su, j'ai presque hésité à reprendre mes études en pharmacie, histoire de te retrouver ! Je te souhaite vraiment de réussir et de t'épanouir dans cette nouvelle fonction. Je te remercie enfin pour ton soutien indéfectible suite à ma maternité, tes conseils et tous les bons moments passés ensemble, au boulot, au resto ou ailleurs. Je veux ensuite remercier Jean-Seb, qui a pris la suite d'Anne pour la dernière partie de ma thèse. Merci pour ta confiance, ton enthousiasme et nos nombreuses discussions. Merci de m'avoir donné deux fois la chance d'assister au CSH.

Je continuerai avec Valérie et Marie-Jeanne. Valérie, une fois Anne partie, tu as été comme ma seconde « Maman du labo », toujours disponible pour un conseil, une discussion. Côté informatique, on se comprenait... Et côté bouffe aussi ! Merci pour tout ce chocolat d'ailleurs... Tu m'auras toi aussi beaucoup appris à la paillasse et en culture, où on aura par ailleurs bien discuté et rigolé. Pour ma part, une amitié est née et j'espère qu'elle perdurera avec les années !

Marie, de qui je pourrai dire « quand je serai grande, je veux être MJP ». En plus de ta rigueur scientifique, ta pédagogie et ta disponibilité malgré un emploi du temps toujours surchargé, ton sens de la déconnade, je retiendrai particulièrement l'affaire Manrique, qui fut un beau moment de solidarité !

Enfin, j'aurais une pensée émue pour Christophe, parti bien trop brutalement et bien trop tôt. C'est en quelques sortes de ta faute si j'en suis là aujourd'hui. Partie en IUP pour revenir rapidement dans le monde du travail, je t'ai croisé sur mon chemin. Excellent enseignant, passionné par sa discipline, motivé comme personne pour faire avancer la recherche, tu m'as

détourné de mes objectifs professionnels pour allumer en moi « cette flamme » qui t'était si chère, m'entraînant dans le tourbillon du doctorat !

Ensuite, je veux remercier évidemment tous les étudiants du labo, qui ont su créer une super ambiance ! Je pense en particulier à notre Anne-So internationale qui est devenue une véritable amie, et qui m'aura aussi appris beaucoup de choses au labo et ailleurs. On aura vraiment bien rigolé toutes ces années, et je ne doute pas qu'on rigolera bien des années encore, sauf si tu me voles mari et enfants ;-). Il y a aussi Théo, évidemment ! Danseur et chanteur hors pair ! Que de discussions nous avons partagées, mûrissant nos projets respectifs autour d'un café, devant un tableau blanc imaginaire ! Tous les trois, nous aurons partagé bières et sushi avec rigueur et persévérance. De très bons souvenirs pour moi et pour Emilie ! Enfin, Laura, la princesse du labo. En plus des nombreux échanges scientifiques, je te remercie pour tes très précieux conseils en termes de mode, quand je me retrouvais face à de délicates situations où un bon jean et un petit bateau ne pouvaient suffire !

Je remercierai aussi les étudiants avec qui j'ai eu la chance de travailler : Marie Large, Ivan (enfin, je veux dire EL Ivan !) et ma chère Patricia. Ce fut à chaque fois une sorte d'angoisse que d'avoir la responsabilité de votre travail mais vous avez tous vraiment assuré et m'avez beaucoup apporté dans mes projets. Et surtout, ce fut avant tout de supers rencontres. Je vous souhaite à tous longue et belle vie en sciences ou ailleurs !

Je terminerai par tous les autres, de l'équipe 2 ou d'ailleurs, Marina, Anfina, Srdana, Elodie, Rémy, Sandrine, Aude, Amandine, Charline, David, Romain, Yvan, Emeline, Cathy, Rémi, Julie. Vous avez tous participé à l'aventure humaine de ma thèse et j'espère que nos chemins se croiseront de nouveau, ici ou là !

Ensuite, il y a tous ceux qui n'étaient pas au labo, mais qui ont vécu et supporté ma thèse. Le premier, le plus touché par le fléau, mon Amour et maintenant mon mari, El Pédro. Lui, qui avait rencontré une serveuse dans un restaurant de la banlieue toulousaine, s'est retrouvé avec une étudiante à charge... Pendant 9 ans !!! Merci pour ton soutien et ta confiance tout au long de ces années ! Je sais que j'ai parfois été bien chi... et que tu n'as pas cillé. J'ai bien peur de l'être encore un peu toujours bien que la thèse soit finie... Et merci d'avoir accepté de trouver chez toi, le soir, en rentrant du travail, des étudiants en train de boire des bières et de pervertir ta fille dès son plus jeune âge par des apéros à répétition !

Je voulais aussi remercier mes parents et ma sœur. Vous m'avez toujours soutenu dans les méandres de mes choix professionnels, me donnant la liberté de faire, de me tromper, de recommencer pour arriver au luxe dans lequel je vis : aimer son travail ! Merci beaucoup d'être toujours là pour prendre soin de moi, pour fêter la moindre occasion autour d'un bon

repas, tous ensemble ! C'est ça qui fait que la vie, parfois dure, reste un plaisir précieux ! Je remercierais enfin tou-te-s mes ami-e-s, rencontré-e-s tout au long de ces années depuis la maternelle (!), Romain et Guéna, Les filles du lycée Saint Sernin, la bande du Lycée hôtelier (Ahhh La bande du Lycée Hôtelier !), les copains de l'hôtellerie (Marie, la team de l'Amphi), les « jeunes » rencontrés à la Fac (en particulier Maya et Tiphaine). Il y a aussi les Cousins, que j'aime tendrement, et ces grands amis qui me suivent depuis mon plus jeune âge : Roger et Liliane, François et Fioune, Michel et Fabienne, Sabine qui auront tous été un soutien dans ma vie de thésarde... et de gourmande !

PRINCIPALES ABRÉVIATIONS UTILISÉES	10
RÉSUMÉ	12
ABSTRACT.....	13
INTRODUCTION GÉNÉRALE	14
PRÉAMBULE	15
CHAPITRE I : LES ORIGINES DE RÉPLICATION	16
I. LA MISE EN PLACE DES ORIGINES DE RÉPLICATION	16
A. La mise en place du complexe de pré-réplication.....	16
B. L'activation des origines de réplication.....	17
C. Prévention de l'activation multiple d'une seule origine	19
D. Spécification des origines de réplication	21
1. Les déterminants génétiques des origines de réplication chez les eucaryotes	21
2. Les déterminants épigénétiques des origines de réplication.....	24
a. La conformation de l'ADN.....	24
b. L'environnement chromatinien	24
c. Réplication et transcription	25
II. ORGANISATION ET RÉGULATION SPATIALE DE L'ACTIVITÉ DES ORIGINES DE RÉPLICATION.....	26
A. Organisation des origines en « réplicon » et flexibilité.	27
B. Le choix des origines actives.....	28
1. L'architecture nucléaire	29
2. Développement et réplication	29
3. Origines de réplication et stress réplcatif.....	30
III. Conclusion	31
CHAPITRE II : LE <i>TIMING</i> DE RÉPLICATION.....	32
I. L'ANALYSE DU <i>TIMING</i> DE RÉPLICATION	32
A. Les foyers de réplication	32
B. L'étude du <i>timing</i> de réplication à l'échelle du génome entier	33
1. Méthodes d'étude.....	33
2. Visualisation du <i>timing</i> de réplication	36
II. LES CARACTÉRISTIQUES GÉNÉTIQUES ET ÉPIGÉNÉTIQUES LIÉES AU <i>TIMING</i> DE RÉPLICATION.....	37
A. La séquence d'ADN	37
B. Relation entre <i>timing</i> de réplication et transcription.....	38
C. Environnement chromatinien	40
1. A l'échelle du nucléosome.....	40
2. A l'échelle du noyau, organisation spatiale	40
III. MÉCANISMES MOLÉCULAIRES DU <i>TIMING</i> DE RÉPLICATION.....	42

A.	Le point de décision du <i>timing</i> de réplication.....	42
B.	Mcm et compétition pour les facteurs limitant de l'activation des origines de réplication.....	42
C.	Rif1 un régulateur conservé du <i>timing</i> de réplication.....	44
1.	Rif1 régule le <i>timing</i> de réplication de la levure à l'homme, en passant par la souris.....	44
2.	Recrutement de Rif1 à la chromatine au niveau de structures en G-quadruplex.	46
IV.	IMPORTANCE BIOLOGIQUE DU <i>TIMING</i> DE REPLICATION	48
A.	Le <i>timing</i> de réplication, « carte d'identité de la cellule »	48
1.	Transmission d'une génération à l'autre	48
2.	Le <i>timing</i> de réplication, signature de l'état de différenciation et du type cellulaire.....	49
B.	<i>Timing</i> de réplication et cancer	50
	RÉSULTATS	52
	OBJECTIFS DE MON TRAVAIL DE THESE.....	53
	CHAPITRE I : DÉCOUVERTE D'UN NOUVEAU RÉGULATEUR DU <i>TIMING</i>	
	DE RÉPLICATION : L'ADN POLYMÉRASE THÊTA.....	54
I.	Les ADN polymerases.....	54
A.	Deux grandes familles d'ADN polymérases	54
1.	Les ADN polymérases répliquatives	55
2.	Les ADN polymérases spécialisées	56
B.	L'ADN polymerase Thêta	56
1.	Structure	56
2.	Fonctions	57
a.	Maintien de l'intégrité du génome et réparation	58
b.	Rôle dans le NHEJ alternatif	58
c.	Autres fonctions décrites	62
3.	Signature dans les cancers	63
II.	Un nouveau rôle pour l'ADN polymérase Thêta dans le <i>timing</i> de réplication	64
III.	Evidences supplémentaires du rôle de Polθ dans le <i>timing</i> de réplication.....	93
A.	Polθ joue un rôle dans l'activation des origines de réplication.....	93
B.	Polθ influence l'environnement et l'organisation chromatinien	94
C.	Polθ régule le <i>timing</i> de réplication dans des cellules humaines normales.	95
IV.	Conclusion et Perspectives	95
A.	Conclusion	95
B.	Perspectives.....	96
1.	Mécanismes d'action de Polθ dans sa fonction régulatrice du <i>timing</i> de réplication.	96
2.	L'impact de la modification du <i>timing</i> de réplication suite à la dérégulation de Polθ.....	97
	CHAPITRE II : MODIFICATION DU <i>TIMING</i> DE RÉPLICATION INDUITE	
	PAR UN STRESS RÉPLICATIF AU CYCLE PRÉCÉDENT	99
I.	La transmission des dommages suite à un stress répliatif	99
A.	Le stress répliatif	99

1.	Définition et causes du stress réplicatif.....	99
a.	Les dommages à l'ADN	99
b.	La dérégulation des facteurs de réplication	100
c.	La séquence nucléotidique	101
d.	Les conflits entre réplication et transcription.....	102
2.	Le cas particulier des sites fragiles communs	103
3.	Réponse cellulaire au stress réplicatif.....	104
B.	La transmission à la génération suivante de dommages induits par le stress réplicatif	105
C.	53BP1 et réparation des cassures double brin	107
1.	Le recrutement de 53BP1 aux CDB.....	107
2.	53BP1 et Rif1 orientent le choix de réparation des CDB.....	108
D.	Problématique	109
II.	Résultats	109
1.	Plan expérimental	109
2.	Le <i>timing</i> de réplication est avancé dans les cellules filles de cellules-mères ayant subi un stress réplicatif	115
a.	Modifications du <i>timing</i> de réplication.....	115
b.	Analyse des domaines affectés	117
c.	Conclusion	118
3.	Mécanismes	118
a.	Corrélation avec les foyers 53BP1.....	118
b.	Recherche de l'implication de l'ADN polymérase Thêta	119
III.	Conclusions et perspectives.....	120
A.	Le <i>timing</i> de réplication, nouveau processus d'adaptation au stress réplicatif ?.....	120
B.	Perspectives.....	122
	CONCLUSIONS ET DISCUSSION GÉNÉRALE.....	125
I.	Polθ régule le <i>timing</i> de réplication	126
II.	Modification du <i>timing</i> de réplication après un stress réplicatif à la génération précédente.	128
III.	Polθ, acteur dans l'adaptation du <i>timing</i> de réplication au stress réplicatif ?.....	129
	ANNEXES	131
I.	ANNEXE I : Classification des gènes situés dans des régions dont le <i>timing</i> de réplication est modifié après déplétion de <i>POLQ</i> dans les cellules RKO.....	132
1.	En fonction des voies de signalisation	132
2.	En fonction des processus biologiques	134
II.	ANNEXE II : Classification des gènes contenus dans les régions tardives dont le <i>timing</i> de réplication est avancé dans les cellules-filles des cellules mères stressées	134
	PROTOCOLES	135
	BIBLIOGRAPHIE	145

PRINCIPALES ABRÉVIATIONS UTILISÉES

ACS : *Autonomous Consensus Sequence*
 ADN : acide désoxyribonucléique
 Alt-NHEJ : NHEJ alternatif
 APC/C : *Anaphase-Promoting Complex/Cyclosome*
 ARN : Acide RiboNucléique
 ARNm : ARN messenger
 ARS : *Autonomously Replicating Sequences*
 ATM : *Ataxia telangiectasia mutated*
 ATP : Adénosine Tri-Phosphate
 ATPase : Adénosine Tri-Phosphate Hydrolase
 ATR : *Ataxia Telangiectasia and Rad3 related*
 BRCA1 : protéine *Breast Cancer 1*
 BrdU : Bromo-déoxyUridine
 CDB cassures double-brin
 Cdk : *Cyclin Dependant Kinase*
 Cdc6 : *Cell division cycle protein 6*
 cdc45 : *Cell Division Cycle 45*
 Cdt1 : *CDC10-Dependant Transcript 1*
 ChIP : Immunoprécipitation de chromatine
 CMG : Cdc45-Mcm-GINS
 c-NHEJ : NHEJ classique
 CTR : *Constant timing Regions*
 Ddk : *Dbf4 Dependant Kinase*
 dNTP : désoxyribonucléotides
 EdU : 5-Ethynyl-2'-deoxyuridine
 FISH : hybridation in situ couplé à de la fluorescence
 G4 : G-quadruplex
 GINS : *Go-Ichi-Ni-San*
 ICG : îlots CpG
 IgG : immunoglobulines
 IP : immuno-précipitation

LINE1 : *Long Interspersed Nuclear Element 1*

Mb : mégabases

Mcm : *Mini Chromosome Maintenance*

MMEJ : *Microhomology Mediated End Joining*

NHEJ : *Non Homologous End Joining* »)

Nt : nucléotides

ODP : *Origin Decision Point*

OGRE : *Origin G-Rich Repeated Element*

Orc : *Origin Recognition Complex*

Pb : paire de bases

PCNA : *Proliferating Cell Nuclear Antigen*

Pol : ADN polymérase

pré-IC : complexe de pré-initiation

pré-RC : complexe de pré-réplication

RFC : *Replication Factor C*

RH : recombinaison homologue

RI : Radiations ionisantes

RPA : *Replication Protein A*

SCF : *Skp1-Cullin-Fbox*

SFC : Sites fragiles communs

SHM : hypermutation somatique

SNS : *Short Nascent Strands*

TDP : *Timing Decision Point*

TTR : *Transition Timing Regions*

UFB : *Ultrafine Bridge*

RÉSUMÉ

La duplication de l'ADN au cours de la phase S est initiée à partir de l'activation de plusieurs dizaines de milliers d'origines de réplication. La mise en place des origines a lieu au cours de la phase G1 sous la forme de complexe de pré-réplication (pré-RC) et leur activation est orchestrée par un programme spatio-temporel. La régulation spatiale détermine les origines qui seront activées et la régulation temporelle, ou *timing* de réplication, détermine le moment de leur activation. En effet, toutes ces origines ne sont pas activées en même temps durant la phase S : certaines origines seront activées en début de phase S, d'autre en milieu, ou d'autre à la fin. Ce programme est établi en tout début de phase G1, au « point de décision du *timing* ». C'est un programme très robuste qui signe l'identité d'une cellule, son état de différenciation et le type cellulaire à laquelle elle appartient. Il a aussi été montré qu'il est altéré dans des situations pathologiques, en particulier le cancer, sans qu'on ne comprenne très bien les raisons mécanistiques. De manière générale, les mécanismes moléculaires qui régulent le *timing* de réplication sont méconnus.

Le premier volet de ma thèse a permis l'identification d'un nouveau régulateur du *timing* de réplication : il s'agit de l'ADN polymérase spécialisée Thêta. Recrutée à la chromatine très tôt en phase G1, elle interagit avec des composants du pré-RC, et régule le recrutement des hélicases réplcatives à la chromatine. Enfin, sa déplétion ou sa surexpression entraîne une modification du *timing* de réplication à l'échelle du génome.

Dans la deuxième partie de ma thèse, j'ai exploré les mécanismes qui régulent ce programme temporel d'activation des origines suite à un stress réplcatif. J'ai identifié un mécanisme de régulation transgénérationnel inédit : la modification du *timing* de réplication de domaines chromosomiques ayant subi un stress réplcatif au cycle cellulaire précédent. Des cellules-filles issues d'une cellule ayant subi des problèmes de réplication dans des domaines fragiles (riches en AT, et donc potentiellement structurés, et pauvres en origines) présentent un *timing* plus précoce de l'activation des origines au niveau de ces domaines. Ce nouveau processus biologique d'adaptation est particulièrement intéressant dans un contexte tumoral de haut stress réplcatif chronique car ce pourrait être un moyen pour la cellule tumorale de survivre à son propre stress réplcatif mais aussi aux thérapies antitumorales qui sont nombreuses à cibler la réplication de l'ADN.

ABSTRACT

DNA duplication in S phase starts from thousands of initiation sites called DNA replication origins. These replication origins are set in G1 as pre-replication complexes (pre-RC) and fired in S phase following a spatio-temporal program of activation. This program determines which origins will be fired and when. Indeed, all the origins are not fired in the same time and we can distinguish early, middle and late replication origins. This temporal regulation is called “replication timing” and is determined at the “timing decision point” (TDP) in early G1. It’s a robust program, which participates to the definition of cell identity, in term of differentiation state or cell type. However, the precise molecular mechanisms involved are poorly understood. Defective timing program has been evidenced in pathological contexts, in particular in cancers, but the mechanisms of this deregulation remain unclear.

In the first part of my PhD, I contributed to the discovery of a new regulator of the origin timing program: the specialized DNA polymerase Theta (Pol θ). Pol θ is loaded onto chromatin in early G1, coimmunoprecipitates with pre-RC components and modulates the recruitment of Mcm helicases at TDP. Moreover, depletion or overexpression of Pol θ modifies the timing of replication at a fraction of chromosomal domains.

The second part of my work aimed at exploring the mechanisms that regulates replication timing after a replicative stress. I identified a totally new transgenerational adaptive mechanism of DNA replication timing regulation: the modification of the timing of origin activation at chromosomal domains that have suffered from a replicative stress during the previous cell cycle. Daughter cells from a cell that has experienced replication stress at particular domains (late replicating domains, AT rich so they can form structured DNA, and poor in origin density) shows advanced origin activation within these regions. This new biological process in response to replicative stress could be of particular interest in the context of cancer since, tumor cells are characterized by high level of intrinsic chronic replicative stress. This new mechanism may favor cancer cell survival despite replication stress, particularly upon treatments with anti-tumor agents that target DNA.

INTRODUCTION GÉNÉRALE

PRÉAMBULE

La duplication du génome a lieu essentiellement au cours de la phase S du cycle cellulaire. Cependant, de plus en plus d'évidences montrent que toutes les phases du cycle cellulaire sont impliquées dans la régulation de la réplication fidèle et complète du génome. En particulier, les origines de réplication, sites à partir desquels la duplication de l'ADN (acide désoxyribonucléique) est initiée au cours de la phase S, sont mises en place pendant la phase G1 du cycle cellulaire. Il y a donc une séparation temporelle entre leur mise en place, appelée « *licensing* », et leur activation, appelée « *firing* », prévenant ainsi des risques de re-réplication.

Le nombre d'origines mises en place est largement supérieur au nombre d'origines qui seront effectivement activées (Cayrou et al., 2011; Depamphilis, 1993), permettant une flexibilité spatiale dans l'initiation de la réplication. En plus de cette régulation spatiale de l'activation des origines de réplication, il existe une régulation temporelle, échelonnant l'activation des origines tout au long de la phase S. Cette flexibilité spatiale et temporelle est importante d'un point de vue biologique car elle permet une adaptation des cellules aux modifications architecturales et structurelles du noyau ainsi qu'à l'activité transcriptionnelle inhérentes à la fonction de la cellule et à sa progression dans une voie de différenciation donnée. Enfin, elle est aussi importante dans la réponse aux dommages à l'ADN et la stabilité du génome, car elle permet une activation des origines optimale pour favoriser la réplication de l'intégralité du génome dans le temps imparti, ou l'activation d'origines de secours pour contourner les multiples obstacles à la progression des fourches de réplication.

Dans cette introduction, je m'attacherai donc à décrire dans un premier temps la mise en place des origines de réplication au cours de la phase G1 du cycle cellulaire et leur activation, et développerai dans une deuxième partie leur régulation temporelle, ou « *timing* de réplication ».

CHAPITRE I : LES ORIGINES DE RÉPLICATION

I. LA MISE EN PLACE DES ORIGINES DE RÉPLICATION

A. La mise en place du complexe de pré-réplication

Les origines de réplication sont mises en place dès la fin de la mitose et au cours de la phase G1 du cycle cellulaire au cours de laquelle le recrutement séquentiel de plusieurs protéines conduit à la formation du complexe de pré-réplication, ou pré-RC, achevée en fin de G1. Cela correspond à l'étape du « *licensing* ». Ensuite, à la transition entre la phase G1 et la phase S, le recrutement de nouveaux acteurs grâce à l'activité des kinases régulatrices Ddk (pour « *Dbf4 Dependant Kinase* ») et Cdk (pour « *Cyclin Dependant Kinase* ») va conduire progressivement à leur « mise à feu », permettant le démarrage proprement-dit de la duplication de l'ADN : c'est le « *firing* » des origines de réplication (Figure 1). Si de nombreuses études ont cherché à décrypter les mécanismes de ces deux étapes de la mise en place des origines de réplication, une récente étude a permis de les reconstituer *in vitro* chez la levure (Yeeles et al., 2015). Tout d'abord, les origines de réplication sont reconnues par le complexe de reconnaissance des origines, « *Origin Recognition Complex* » (Orc). C'est un hexamère constitué de 6 sous-unités, Orc1 à 6, dont cinq d'entre elles, Orc1-Orc5, présentent une activité Adénosine Tri-Phosphate Hydrolase (ATPase) associée (Neuwald et al., 1999; Speck et al., 2005). Les protéines Orc lié à l'ADN servent ensuite de plateforme pour le recrutement de Cdc6 (« *Cell division cycle protein 6* »), une autre ATPase, formant avec Orc un anneau autour de l'ADN (Sun et al., 2012). Cette formation du complexe Orc-Cdc6-ADN requiert de l'ATP (Speck et al., 2005). A son tour, elle permettra le recrutement ATP-dépendant de Cdt1 (« *CDC10-Dependant Transcript 1* ») et des hélicases réplcatives Mcm2-7 (*Mini Chromosome Maintenance 2-7*), autres AAA+ ATPases, sous leur forme inactive. Au total, deux hexamères Mcm2-7 seront recrutés séquentiellement au niveau d'une origine, formant ainsi un double-hexamère qui encercle l'ADN d'un même côté du complexe Orc-Cdc6 (Gambus et al., 2011; Remus et al., 2009; Tica et al., 2015). Chez la levure ou la souris, il a été montré que le recrutement de Cdt1-Mcm2-7 se fait sous la forme du recrutement d'un complexe préalablement formé dans le nucleoplasme (Tanaka and Diffley, 2002). En revanche, rien de tel n'a été observé dans les cellules humaines et on ne peut exclure que le complexe Orc-Cdc6-ADN recrute d'abord cdt1, qui recrutera à son tour les deux hexamères Mcm2-7. Une fois les Mcm chargés au niveau des origines, le complexe de pré-réplication (pré-RC) est formé, et l'étape de *licensing* terminée.

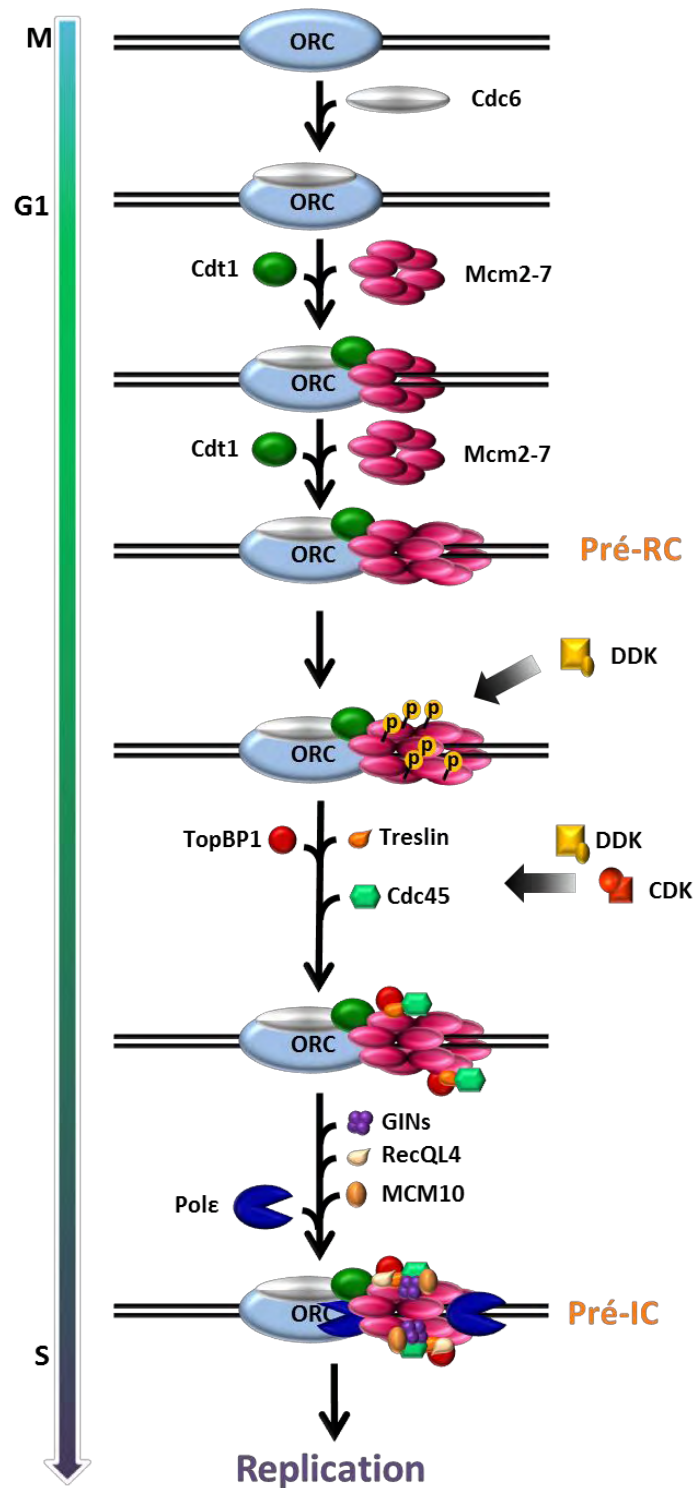


Figure 1 : Mise en place des origines de réplication. Adaptation de (Fragkos et al., 2015). Dès la fin de la mitose (M), et au cours de la phase G1, les complexes de pré-réplication (pré-RC) sont mis en place. C'est le *licensing*. Ensuite, à la transition G1/S, l'activation des kinases DDK et CDK induit l'activation des pré-RC pour former les complexes de pré-initiation (pré-IC) d'où démarrera la réplication en phase S.

B. L'activation des origines de réplication

A la transition G1/S, l'activité des kinases Ddk et Cdk induit la formation du complexe de pré-initiation (pré-IC) par la phosphorylation et le recrutement de nouveaux facteurs au

niveau des pré-RC. Les double-hexamères Mcm inactifs vont ainsi être convertis en un complexe Cdc45-Mcm-GINS (CMG) actif qui constituera l'hélicase répliquative. Les différentes études semblent montrer que les mécanismes de mise en place du pré-IC diffèrent un peu selon les organismes. D'une part, chez la levure, les DDK, suite à la phosphorylation des Mcm, induisent le recrutement de cdc45 (« *Cell Division Cycle 45* ») et de Sld3/7. Ensuite, DDK et CDK orchestrent le recrutement de Sld2, Dpb11, du complexe GINS (pour « *Go-Ichi-Ni-San* », 5,1,2,3 en Japonais), de l'ADN polymérase epsilon (Pol ϵ) et enfin de Mcm10. Le complexe CMG est ainsi formé et permet l'ouverture de la double-hélice d'ADN et le recrutement de la Protéine de Réplication A (RPA, pour « *Replication Protein A* ») sur l'ADN simple-brin. Chez le xénope, la formation du pré-IC commence par l'association cdk2-dépendante de la Treslin, analogue de sld3, avec l'ADN-topoisomérase TopBP1 (pour « *DNA Topoisomerase II Binding Protein 1* »), analogue de Dpb11, au niveau des origines, et permet le recrutement de cdc45 (Kumagai et al., 2010). Dans le modèle humain des cellules Hela, il semblerait cependant que TopBP1 ne soit pas indispensable au recrutement de Cdc45 (Im et al., 2009). Ensuite, les recrutements de l'hélicase RecQL4 (« *ATP-dependant DNA helicase Q4* », homologue de Sld2 chez la levure), du complexe GINS, de Pol ϵ et Mcm10 permettront de finaliser la formation et l'activation du CMG et enfin le recrutement de RPA et de l'ADN polymérase α (Im et al., 2009; Matsuno et al., 2006). Avec le recrutement des autres facteurs constitutifs du réplisome, dont l'ADN polymérase δ , le facteur de processivité PCNA (« *Proliferating Cell Nuclear Antigen* ») et son *clamp loader* associé RFC (« *Replication Factor C* »), la réplication peut alors commencer (Figure 2). C'est l'étape d'élongation, assurée par la progression d'une fourche de réplication bidirectionnelle. Les hélicases répliquatives sont constituées de Cdc45 associé à Mcm2-7 et au GINS, formant le complexe CMG. Situé en aval du réplisome au sein d'un grand complexe de progression du réplisome (complexe RPC pour « *Replisome Progression Complex* »), il déroule l'ADN, rendant ainsi la matrice d'ADN accessible aux ADN polymérases répliquatives. Associées à leur facteur de processivité PCNA, elles vont dupliquer l'ADN sur chaque brin à partir des amorces d'ARN (Acide RiboNucléique) synthétisées par Pol α (Pour revue (Garg and Burgers, 2005)). La double hélice d'ADN étant constituée de deux brins antiparallèles, et les ADN polymérases répliquatives ne pouvant synthétiser l'ADN que dans le sens 5'-3', un brin sera répliqué de façon continue, le brin direct, et l'autre de façon discontinue, le brin indirect. La duplication de ce brin indirect consiste donc en une succession de fragment d'ADN, d'environ 200pb chez les Eucaryotes, appelés les fragments d'Okasaki (Okazaki et al., 1968). La synthèse de ces fragments nécessite la synthèse d'une amorce ARN :ADN par Pol α , avant que l'ADN

polymérase répliquative ne prenne le relais pour synthétiser l'ADN jusqu'à rencontrer la prochaine amorce synthétisée par Pol α . Elle va alors déplacer le brin d'ARN pour le remplacer par de l'ADN. L'extrémité flottante qui en résulte est coupée par l'endonucléase Fen1. La ligation des deux fragments d'ADN est catalysée par la ligase I.

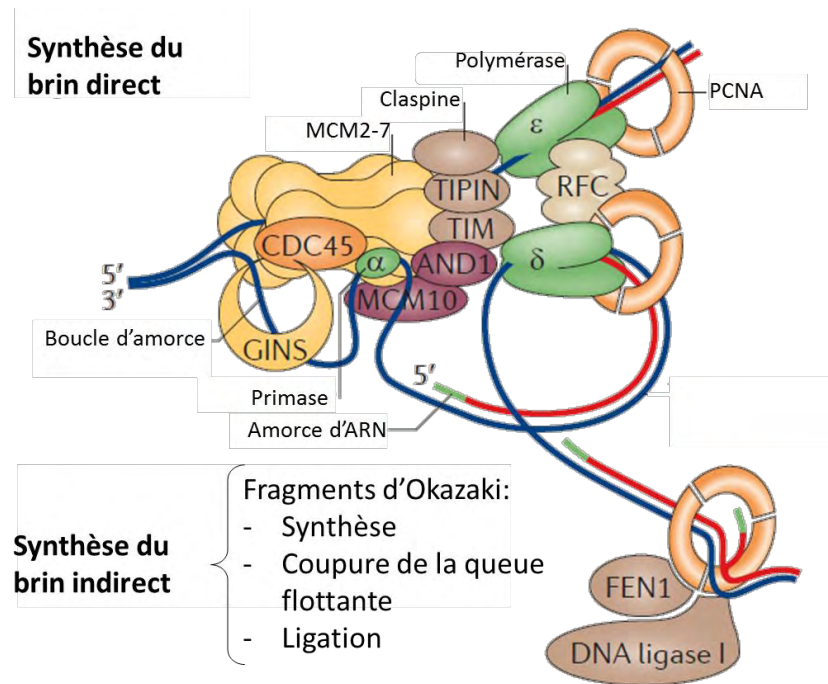


Figure 2 : Réplisome eucaryote. Tiré de (Alabert and Groth, 2012)

C. Prévention de l'activation multiple d'une seule origine

S'il est important qu'à chaque cycle cellulaire, l'information génétique soit copiée fidèlement et dans son intégralité, il est aussi important qu'elle ne soit copiée qu'une seule fois, au risque d'amplification, délétion ou aneuploïdie. La séparation dans le temps du « *licensing* » et du « *firing* » des origines joue un rôle primordial pour éviter la re-réplication, c'est-à-dire l'initiation multiple à partir de la même origine. Il faut donc que l'activation des Mcm ne puissent pas avoir lieu au cours de la phase G1 du cycle cellulaire, ni que les complexes de pré-réplication puissent se reformer après l'activation d'une origine en phase S.

Ce décalage temporel est tout d'abord assuré par le jeu de deux enzymes, les CDK, et l'ubiquitine-ligase *Anaphase-Promoting Complex/Cyclosome*, APC/C (pour revue, (Siddiqui et al., 2013)). En fin de mitose, cette dernière catalyse l'ubiquitination des cyclines spécifiques des phases S et G2-M, cycline A et B respectivement, induisant ainsi leur dégradation par le protéasome (Sudakin et al., 1995). De ce fait, l'activité de leur CDK associée est abolie. Comme APC/C reste active au cours de la phase G1, cela crée, au cours

de cette phase, une fenêtre dépourvue d'activité des CDK, et l'activation des origines de réplication qui sont en train de se mettre en place est inhibée. A l'inverse, la forte activité des Cdk pendant les phases S, G2 et M inhibent l'activité d'APC/C, ainsi que la mise en place de nouvelles origines. Il y a donc une inter-régulation entre Cdk et APC/C.

Ensuite, la régulation du *licensing* au cours du cycle cellulaire est assurée par la régulation des acteurs du pré-RC eux-même. Tout d'abord, le recrutement de l'hexamère Orc aux origines de réplication est régulé par le relargage d'une de ses sous-unités de la chromatine, Orc1. La phosphorylation d'Orc1 par le complexe Cycline A-Cdk2 induit sa poly-ubiquitination par le complexe ubiquitine-ligase *Skp1-Cullin-Fbox* (SCF). Orc1 est alors adressé au protéasome puis dégradé (Li and DePamphilis, 2002; Méndez et al., 2002). Cette déstabilisation d'Orc1 empêche donc la formation de nouveaux complexes Orc à la chromatine en dehors de la phase G1. D'autre part, dans les cellules humaines, Cdc6 est en grande partie dégradé assez tôt en phase G1 suite à l'activation du complexe APC/C et de son cofacteur cdh1 (Clijsters et al., 2013). Par ailleurs, il a été montré que Cdc6 pouvait être phosphorylé dès l'entrée en phase S par cyclin A-cdk2, induisant sa délocalisation dans le cytoplasme (Petersen et al., 1999), évitant tout nouveau recrutement au niveau des complexes ORC libérés au fur et à mesure de la réplication. Le troisième élément à être recruté au cours de la mise en place du pré-RC est Cdt1. L'inhibition de son recrutement à la chromatine en phase S se fait à différents niveaux. Il est tout d'abord la cible de deux ubiquitine-ligases qui entraînent sa déstabilisation. Au cours de la phase S, l'interaction de Cdt1 avec PCNA une fois l'origine activée induit sa dégradation par l'ubiquitine-ligase Cul4/Ddb1 (« *Cullin4/DNA Damage Binding protein 1* »), appelée aussi CRL4-Cdt2 (« *Cullin Ring E3 Ligase-Cdc10 Dependant Transcript 2* »). De plus, chez l'Homme, Cdt1 est lui aussi phosphorylé par Cycline A-Cdk2, induisant sa dégradation par une voie dépendant de SCF^{Skp2} en phases S et G2. Si la régulation par CRL4-Cdt2 est essentielle dans la prévention de la re-réplication, celle orchestrée par SCF^{Skp2} semble jouer un rôle moins important (Arias and Walter, 2006; Nishitani et al., 2006; Senga et al., 2006). Par ailleurs, un acteur très important dans la prévention de la re-réplication par la régulation de l'activité de Cdt1 est la géminine. La géminine, en se liant directement à Cdt1, la séquestre et empêche ainsi son interaction avec les hélicases Mcm et donc leur recrutement à la chromatine. Etant un substrat de l'ubiquitine-ligase APC/C, elle est dégradée en fin de mitose, permettant la formation du pré-RC au cours de la phase G1 (Clijsters et al., 2013; Li and Blow, 2005; McGarry and Kirschner, 1998; Wohlschlegel et al., 2000). Dès l'activation des kinases réplicatives, APC/C est réprimée et la géminine à nouveau exprimée. Chez la levure, il a été aussi montré que la phosphorylation des

Mcm par la CDK cdc28, associée à la cycline régulatrice Clb (clb/cdc28) activée à l'entrée en phase S, induit l'export des Mcm hors du noyau (Labib et al., 1999; Nguyen et al., 2000). La conservation chez les Eucaryotes des sites de phosphorylation de Mcm4 par les CDK suggère que ce mécanisme pourrait être conservé chez les Eucaryotes supérieurs et serait en accord avec l'observation du relargage des Mcm de la chromatine pendant la phase S (Moses et al., 2007; Méndez and Stillman, 2000; Prasanth et al., 2004).

Enfin, dans un contexte favorable à la re-réplication, comme la surexpression de Cdt1 par exemple, l'activation du point de contrôle de phase S dépendant d'ATR (« *Ataxia Telangiectasia and Rad3 related* ») inhibe la re-réplication par la phosphorylation directe de cibles telles RPA ou Mcm2. Elle induit aussi la déphosphorylation de la protéine du Rétinoblastome (Rb) (Liu et al., 2007), dont la phosphorylation par le couple CyclineD-Cdk4 en fin de G1 induit l'expression de la Cycline E et de la Cycline A permettant l'activation des origines de réplication (Bartek et al., 1996).

D. Spécification des origines de réplication

1. Les déterminants génétiques des origines de réplication chez les eucaryotes

Chez la levure *Saccharomyces cerevisiae*, la réplication commence au niveau de régions riches en AT, appelées séquences ARS pour *Autonomously Replicating Sequences*. Ces séquences contiennent une séquence consensus (ACS pour *Autonomous Consensus Sequence*) reconnue par le complexe ORC (*Origin Binding Complex*), premier acteur dans la mise en place des origines de réplication. (Pour revue (Dhar et al., 2012)).

En revanche, une telle séquence consensus n'existe pas chez les eucaryotes supérieurs. En effet, des études *in vitro* dans deux modèles différents ont montré que le complexe ORC, qui reconnaît les origines de réplication, présentait peu, voire aucune spécificité de séquences (Remus et al., 2004; Vashee et al., 2003). Les travaux de Remus suggéraient en revanche que la topologie de l'ADN influençait l'efficacité de liaison d'ORC à la chromatine, l'ADN superenroulé augmentant l'affinité de ORC pour son substrat sans que la spécificité de séquence n'entre en jeu. Plus récemment, un ensemble de travaux suggère que les origines de réplication sont associées à la présence d'ilôts CpG (ICG) (Besnard et al., 2012; Cadoret et al., 2008; Cayrou et al., 2011, 2012). Dans ces trois études, les auteurs ont déterminé les sites d'initiation de la réplication par la méthode dite des « *Short Nascent Strands* » (SNS), qui consiste à identifier les séquences des brins naissants, c'est-à-dire les brins d'ADN néosynthétisés contenant une amorce d'ARN. De façon à ne pas prendre en considération les

fragments d'Ogazaki, et tout en imposant une précision de séquence, les brins ARN-ADN identifiés ont préalablement été sélectionnés en fonction de leur taille (entre 0,5 et 2kb). Ils ont pu ainsi montrer dans plusieurs modèles que chez les eucaryotes supérieurs, une grande partie des sites d'initiation de la réplication est associée à la proximité de régions riches en GC. De plus, Cayrou *et al.* ont montré chez la souris, qu'au niveau du site d'initiation, il existe une asymétrie de séquence de nucléotidique spécifique. En 5', les T et G sont surreprésentés, alors qu'en 3' ce sont les A et C qui sont les plus abondants (Cayrou *et al.*, 2011). Ces régions riches en GC définissent « éléments répétés riches en G associés aux origines », ou OGRE pour « *Origin G-Rich Repeated Element* », situés à quelques centaines de nucléotides du site d'initiation proprement dit (Figure 3A et B). Une grande partie des ICG associés aux origines contiennent deux OGRE, un sur chaque brin d'ADN, disposés en orientation inverse. Ainsi, il y a, à leur niveau deux sites d'initiation de la réplication situés en amont ou en aval de l'OGRE, expliquant le profil bimodal de brins naissants observé pour certaines origines associées au ICG. Inversement, dans le cas où les OGRE ne sont pas inclus dans un ICG, les auteurs n'ont détecté qu'un seul site d'initiation. Enfin, une partie de ces OGRE (50%) contiennent des séquences capables de former des G-quadruplex (G4) (Figure 3C), structures secondaires d'ADN formés par la superposition de plateaux qui se forment entre les guanines d'une séquence riche en G. Réciproquement, plus de 80% des G4 potentiels sont situés dans des OGRE (Cayrou *et al.*, 2012). Cette corrélation entre G4 et origines de réplication est également décrite chez l'humain dans les travaux de Besnard *et al.*, ou plus récemment chez la drosophile (Comoglio *et al.*, 2015). Enfin, une étude réalisée dans les cellules de poulet DT40 vient appuyer, si ce n'est démontrer, le rôle des G4 dans la spécification et l'activation des origines de réplication (Valton *et al.*, 2014). En effet, par des manipulations génétiques au niveau de deux origines caractérisées dans ce modèle, les auteurs de l'étude montrent la nécessité de la présence d'un ou plusieurs G4 associés à ces origines pour qu'elles soient actives. De plus, leur efficacité est corrélée à la capacité à la séquence associée à l'origine de former un G4. En outre, l'orientation du G4 détermine précisément la position du site d'initiation. En revanche, pour l'une des deux origines, les auteurs montrent que le G4 seul ne suffit pas à son activation, mais qu'il est nécessaire qu'il soit associé à une séquence de 200 paires de base (bp) et concluent ainsi que le G4 peut être nécessaire mais pas suffisant pour l'activation d'une origine. Enfin, il est intéressant de noter que ces résultats ont été mis en évidence selon deux approches expérimentales indépendantes : l'identification des SNS, mais aussi l'induction d'une modification dans le *timing* de réplication d'une séquence donnée, s'affranchissant ainsi des biais potentiels de la méthode des brins naissants. Enfin,

déterminées par la présence de G4 à proximité du site d'initiation de la réplication. Cela implique donc aussi l'existence d'origines qui ne répondent pas à ces critères de séquence d'ADN. Cette absence de séquence consensus stricte pour les origines de réplication chez les métazoaires est aussi peut-être un moyen d'adaptation des cellules à leur environnement et leur identité. En effet, la diversité des sites potentiels d'initiation de la réplication permettrait une flexibilité permettant de garantir au mieux l'activation correcte des origines de réplication en toutes circonstances.

2. Les déterminants épigénétiques des origines de réplication

L'absence de séquence consensus universelle chez les métazoaires, ainsi que les premières observations faites par Remus *et al.* suggèrent donc que la séquence à elle seule ne suffit pas et qu'il y existe d'autres déterminants pour définir les origines de réplication. Ainsi, la conformation de l'ADN, l'environnement chromatinien et l'activité transcriptionnelle ont potentiellement un rôle dans la détermination des origines de réplication.

a. La conformation de l'ADN

Remus *et al.* avaient donc observé que l'affinité de ORC *in vitro*, augmentait lorsque l'ADN était super-enroulé, posant ainsi la question d'un rôle de la conformation de l'ADN dans la définition des origines de réplication (Remus et al., 2004) De la même façon, une récente étude chez la drosophile conclut que c'est la conformation de l'ADN qui signe une origine de réplication, chez la drosophile, mais aussi chez l'Homme (Comoglio et al., 2015). En effet, au niveau des origines de réplication, la diminution de la torsion de la double hélice d'ADN ou bien l'élargissement du petit sillon permettent d'augmenter localement la flexibilité de l'ADN et l'ouverture de la double-hélice.

b. L'environnement chromatinien

Plusieurs études ont montré que chez la levure et les eucaryotes supérieurs, les sites de liaison du complexe ORC sont dépourvus en nucléosome (Berbenetz et al., 2010; Besnard et al., 2012; Eaton et al., 2010; MacAlpine et al., 2010), ORC étant lui-même à l'origine de cette exclusion nucléosomique. Il a été suggéré que les sites de liaison à ORC sont en fait occupés par des nucléosomes labiles, enrichis en histones H3.3/H2AZ (Lombrana et al., 2013), facilement déplacés par ORC en G1. En revanche, de façon intéressante, chez la drosophile ou les mammifères, les sites d'initiation de la réplication sont fortement associés à la présence d'un nucléosome (Cayrou et al., 2012; Comoglio et al., 2015; Lombrana et al., 2013). Chez la

drosophile, ces nucléosomes présentent plusieurs marques épigénétiques dont l'Histone 4 Lysine 20 méthylée (H4K20me1) qui favorise le chargement du pré-RC à la chromatine (Tardat et al., 2010) et dont la présence est fortement associée aux origines de réplication chez l'humain (Picard et al., 2014). Les travaux de Lombrana *et al.* montrent que ce nucléosome est déplacé au moment de l'initiation de la réplication. Les deux composantes des origines, le site de liaison du pré-RC et le site d'initiation proprement dit, seraient donc différents de par leur occupation nucléosomique : le site de liaison à ORC serait associé à un nucléosome labile, qui le protégerait tant que le pré-RC n'est pas formé, et le site d'initiation lui-même associé à un nucléosome stable, qui ne serait déplacé qu'au moment de l'initiation de la réplication.

D'autres marques d'histones sont associées aux origines et à leur activation sans qu'il y ait de marque particulière universelle des origines de réplication eucaryotes. Les plus fréquentes sont l'acétylation des histones, qui marque une chromatine décondensée ou la méthylation des histones H3 et H4. Enfin, on peut trouver des remodeleurs de chromatine tel l'*Histone acétyltransférase Binding to ORC1* (HBO1) (Miotto and Struhl, 2008). Une fois de plus, cette diversité dans les marques épigénétiques associées aux origines de réplication pourraient être révélateur d'une diversité nécessaire des origines de réplication dans un génome complexe permettant l'adaptation de la cellule aux contraintes liées à son environnement ou son identité.

c. Réplication et transcription

L'ensemble des études récentes réalisées à l'échelle du génome pour identifier les origines de réplication ont mis en évidence un lien entre transcription et origines de réplication. En effet, les premiers résultats obtenus dans deux études menées dans des cellules humaines d'une part (Cadoret et al., 2008) et de souris d'autre part (Sequeira-Mendes et al., 2009) montrent que les régions riches en gènes sont aussi riches en origines, et qu'une grande partie des origines se situent au niveau de promoteurs de la transcription. Dans ce cas-là, les sites d'initiation de la réplication coïncident avec les sites d'initiation de la transcription (en anglais « *Transcription Starting Sites* » ou TSS). Ces résultats ont été retrouvés par Cayrou *et al.*, qui montrent que cette forte représentation des origines aux TSS chez la souris est en fait une conséquence de leur association aux îlots CpG. Chez la drosophile, parmi les origines activées situées au sein de régions riches en G et en C (CGR), les événements qui prédisent le mieux cette activité sont la transcription de gènes, la méthylation de la lysine 36 de l'histone H3 (H3K36me1), qui est associée à la régulation de la transcription (pour revue, (Wagner and

Carpenter, 2012)) et la liaison de l'ARN polymérase II (Comoglio et al., 2015). Enfin, il semble que des études en cours montrent que certaines origines de réplication se situent au sein de *R-loops*, ces structures tri-caténaires correspondant à l'hybridation d'un ARN sur un des deux brins d'ADN (Chedin, communication au Cold Spring Harbor Laboratory meeting, septembre 2015). Ces travaux sont basés sur l'identification des *R-loops* chez l'Humain ou la souris par immunoprécipitation des séquences double-brin ADN-ARN (ou DRIP pour « *DNA-RNA ImmunoPrecipitation* ») suivi de leur séquençage (DRIPseq). Les résultats obtenus suggèrent fortement que les origines de réplication activées en début de phase S se situent au niveau de *R-loops*. En effet, les sites identifiés par DRIPseq correspondent aux origines précoces identifiées par l'équipe de David Gilbert dans les cellules K562. En outre, les *R-loops* sont associées à des marques d'histones impliquées dans la spécification des origines (H4K20me1, ou H3K79me2), et *in vitro*, ORC1 présente une forte affinité avec les structures double-brin hybrides ARN :ADN. Ces résultats font écho à un article paru en 2015 qui posait justement la question de la présence des origines de réplication au niveau des *R-loops* (Lombrana et al., 2015). Les auteurs proposaient alors que les *R-loops* pouvaient permettre la formation de G-quadruplex, d'ADN ou hybrides ADN :ARN, en dehors de la phase S et ainsi servir au recrutement du complexe ORC, de par l'affinité de sa sous-unité ORC1 pour les structures en G4 (Hoshina et al., 2013).

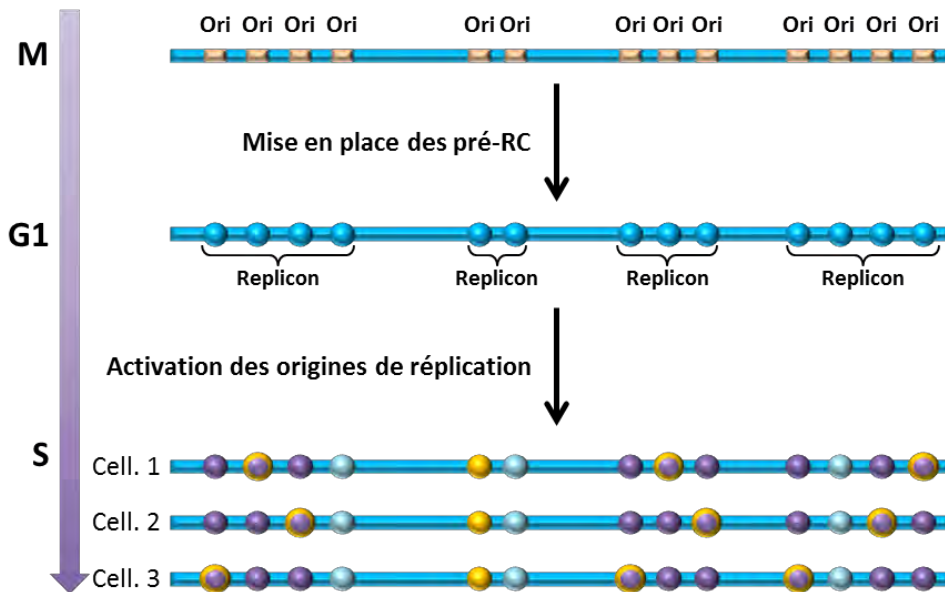
En revanche, dans un contexte de forte activité transcriptionnelle, le recrutement du complexe de pré-réplication et les événements d'initiation de la réplication peuvent être perturbés (Martin et al., 2011).

II. ORGANISATION ET RÉGULATION SPATIALE DE L'ACTIVITÉ DES ORIGINES DE RÉPLICATION

Dès le début des années 1990, DePamphilis proposait un modèle selon lequel il existait plus d'origines potentielles mises en place dans une cellule que d'origines réellement activées. C'est le modèle « Jésuite » inspiré du dicton « Il y a beaucoup d'appelés, mais peu d'élus » (DePamphilis, 1993). Et en effet, les récentes études à l'échelle du génome entier ont montré que chez l'homme ou la souris, plusieurs dizaines de milliers d'origines sont activées dans une cellule au cours de la phase S, alors que le nombre d'origines potentielles est beaucoup plus important (Cayrou et al., 2011; Besnard et al., 2012; Mesner et al., 2011). Cela suggère donc une régulation spatiale dans l'activation des origines de réplication.

A. Organisation des origines en « réplicon » et flexibilité.

A.



B.

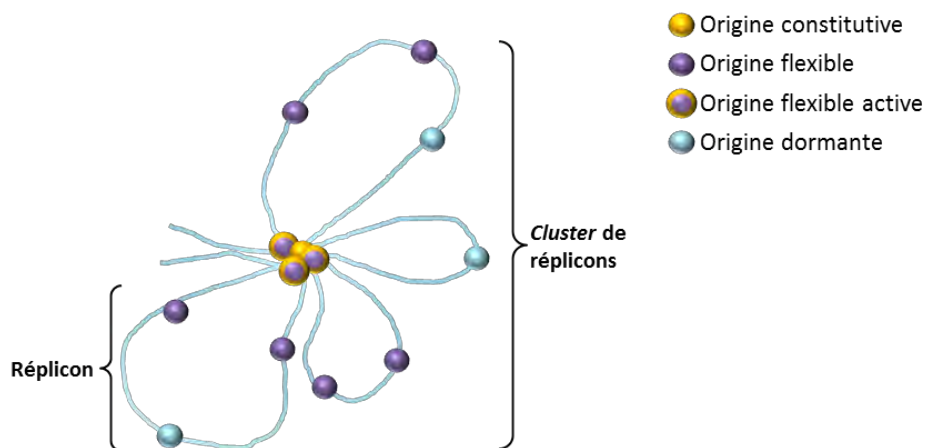


Figure 4 : Organisation spatiale des origines de réplication. (A) Les origines de réplication sont organisées en groupes d'origines appelés réplicon. Au sein d'un réplicon, toutes les origines ne sont pas activées, distinguant ainsi trois classes d'origines : les origines constitutives, les origines flexibles et les origines dormantes. (B) Les réplicons sont eux même regroupés en clusters, les origines activées étant regroupées dans un même espace, à la base de boucles de chromatine. Adaptation de (Cayrou et al., 2011).

Les travaux de Cayrou *et al.* montrent par la méthode d'étirement des fibres d'ADN qu'une origine sur quatre ou cinq est activée au sein d'un réplicon dans un cycle donné, pour une cellule donnée (Figure 4). Cependant, cette activation d'une partie des origines ne se fait pas de façon stochastique à l'échelle du génome. Les origines sont regroupées en réplicons, au sein desquels elles sont fonctionnellement liées, l'activation d'une origine inhibant celle des autres (Lebofsky et al., 2007). Le choix de l'origine activée serait donc stochastique à l'échelle du réplicon. Cette flexibilité dans l'activation des origines permettrait de maximiser

les chances, pour un réplicon, d'être activé au cours du cycle, et permettre ainsi une adaptation de l'activation des origines de réplication aux contraintes auxquelles sont soumises les cellules (programme transcriptionnel, développement, stimulation de croissance, stress réplicatif...) (Gilbert, 2007).

Cependant, le choix de l'origine activée n'est pas complètement aléatoire au sein d'un réplicon, et la fréquence d'activation varie d'une origine à l'autre (Cayrou et al., 2011; Heichinger et al., 2006).

On distingue ainsi plusieurs classes d'origines de réplication : les origines constitutives, les origines flexibles et les origines dormantes (Figure 4). Les origines constitutives sont activées à chaque cycle cellulaire pour chaque cellule d'une population donnée. En revanche, l'activation des origines flexibles paraît plus aléatoire. Elles sont activées différemment d'une cellule à l'autre au sein d'une population donnée et jouent donc un rôle prépondérant dans l'adaptation ponctuelle de l'initiation de la réplication au contexte cellulaire. Enfin, les origines dormantes ne sont pas activées dans le cadre de conditions normales de réplication, mais seront activées en cas de problème de progression de la réplication et de fourches bloquées. Elles jouent donc un rôle fondamental dans la réponse au stress réplicatif (Ibarra et al., 2008).

Un deuxième niveau d'organisation spatiale consisterait au regroupement des réplicons eux même en *cluster*, définissant des domaines de réplication (Figure 4). Au sein de ces domaines l'activation des différents réplicons qui le composent se fait en même temps, et au même endroit, formant les foyers de réplication visualisés en microscopie (Marheineke and Hyrien, 2004). L'ADN serait organisé sous la forme de grandes boucles dont les bases seraient regroupées ensemble, facilitant le recrutement des facteurs d'initiation en concentrant les besoins sur un site restreint, et seraient le siège des origines de réplication activées en phase S. Ce regroupement des origines se ferait grâce aux cohésines qui interagissent avec les MCM d'une part et stabilisent la formation des boucles d'ADN d'autre part (Guillou et al., 2011).

B. Le choix des origines actives

Des travaux réalisés sur le gène DHFR chez le hamster indiquent que le choix des origines qui seront activées en phase S se fait pendant la phase G1, au point de décision des origines (*Origin Decision Point*, ou ODP) (Wu and Gilbert, 1996).

Les facteurs influençant le choix des origines activées en phase S parmi celles mises en place en phase G1 sont multiples. Ces facteurs sont en partie aussi ceux qui influencent le

« *licensing* », comme décrit précédemment. Tout d'abord, la présence de G-quadruplex à proximité de l'origine constitue un facteur *Cis* déterminant de l'activité des origines de réplication, et les origines les plus efficaces y sont associées. Ainsi, les origines constitutives, partagées par plusieurs lignées cellulaires sont associées à leur présence.

Mais le choix de l'origine active dépend aussi, et certainement surtout, de l'environnement chromatinien. Par exemple, nous avons vu que la transcription, la méthylation des îlots CpG ou la modification d'histones influencent, positivement ou négativement, l'activité des origines de réplication. D'autres facteurs sont importants dans la régulation spatiale des origines de réplication.

1. L'architecture nucléaire

Au contraire du *licensing*, le *firing* des origines de réplication nécessite la formation de la membrane nucléaire, déterminant le choix des origines activées (Sheehan et al., 1988). En outre, plusieurs études montrent que plusieurs composants du pré-RC sont recrutés au niveau de la membrane nucléaire (pour revue, (Wilson and Coverley, 2013)). Et *in vitro*, l'initiation de la réplication a lieu au niveau de sites d'attachement de l'ADN à la matrice nucléaire, à la base des boucles de réplication qui constituent les domaines de réplication. Cette localisation des origines activées au niveau de la matrice aurait lieu en fin de phase G1 (Radichev et al., 2005), certainement suite à l'ODP. Cela est cohérent avec le fait que l'ADN néosynthétisé se trouve à la matrice nucléaire.

2. Développement et réplication

Le stade de développement influence grandement l'organisation spatiale des origines de réplication. Dans des stades précoces du développement du xénope ou de la drosophile, les origines activées sont nombreuses, et distribuées sur l'ensemble des chromosomes à intervalles réguliers. La distance moyenne interorigine est alors beaucoup plus courte que dans des stades plus avancés du développement (Hyrien et al., 1995; Sasaki et al., 1999). De plus, cette redistribution des origines de réplication au cours du développement semble liée à la transcription, puisqu'elle intervient au moment où la transcription reprend. En effet, au cours des premiers stades du développement, ce sont les ARN messagers hérités de la mère qui servent à la production des protéines, et aucune transcription n'est donc nécessaire. La réplication commencerait alors à partir d'une multitude de sites, permettant une réduction drastique du temps de la phase S (quelques vingtaines de minutes chez le xénope par exemple). Par ailleurs, l'architecture nucléaire est un autre facteur pouvant influencer

l'organisation spatiale des origines de réplication au cours du développement. En effet, la taille des boucles d'ADN corrèle avec celle des réplicons (Buongiorno-Nardelli et al., 1982; Courbet et al., 2008; Lemaitre et al., 2006). L'organisation du noyau en boucles de chromatine et la définition des réplicons ont lieu pendant la mitose, de façon topoiosomérase II dépendante et corrèle avec le recrutement de Orc à la chromatine. Cette organisation sub-nucléaire est modifiée au cours du développement ou de la reprogrammation cellulaire. En effet, la mise en présence de noyaux de cellules différenciées avec des extraits mitotiques d'œufs de xénope induit une diminution de la taille des boucles de chromatine, reflétant une réorganisation des origines de réplication et une reprogrammation de ces noyaux (Lemaitre et al., 2006). Ainsi la distribution des origines activées participe à la définition de l'identité de la cellule et ses capacités de pluripotence.

3. Origines de réplication et stress réplicatif

Comme nous le verrons plus tard, le stress réplicatif consiste en une accumulation de ralentissements ou d'arrêts de la progression des fourches de réplication. Cette perturbation générale de la progression des fourches va influencer l'activation des origines de réplication de deux façons : l'activation des origines dormantes adjacentes, et l'inhibition des origines de réplication n'ayant pas encore été activées, (Pour revue (Yekezare et al., 2013)). D'un côté, l'activation des origines dormantes permet à la cellule de compenser l'arrêt de la fourche en initiant la réplication sur un site situé en aval de ce blocage de fourche, et assurer ainsi la réplication complète des régions problématiques. D'un autre côté, l'inhibition des origines distales permet d'éviter la formation de nouvelles fourches de réplication dans un contexte de stress réplicatif élevé, et donc non favorable à leur bonne progression. L'inhibition des origines qui auraient dû être répliquées ultérieurement est médiée par le point de contrôle (*Checkpoint*) de phase S, qui est activé en cas de stress réplicatif, induit un arrêt de la progression du cycle cellulaire, et protège la fourche bloquée le temps de gérer les problèmes de réplication. Ainsi, deux des cibles de ce point de contrôle sont les DDK et indirectement les CDK, dont l'inhibition empêche tout *firing*. En revanche, le mécanisme par lequel il active les origines adjacentes est peu connu.

Enfin, il est intéressant de noter que ce *checkpoint*, et en particulier sa kinase effectrice Chk1 (Chk1), régule aussi le nombre d'origines activées en absence de stress réplicatif externe élevé. En effet, sa délétion dans un modèle humain ou chez le poulet induit une augmentation de l'activation des origines de réplication sur l'ensemble du génome (Maya-Mendoza et al., 2007).

III. Conclusion

En conclusion de cette première partie, les Eucaryotes, et en particuliers les Eucaryotes supérieurs, présentent une grande flexibilité dans la détermination des origines de réplication, que ce soit en termes de site de recrutement des protéines du pré-RC, ou en termes de choix des origines activées. Cela procure à la cellule une grande faculté d'adaptation aux multiples contextes auxquels elle peut avoir à faire face : développement et différenciation, micro-environnement, stress répliatif endogène ou exogène... La réplication étant une étape cruciale dans la division d'une cellule et dans le maintien de l'intégrité du génome, elle ne peut souffrir d'une trop grande rigidité qui risquerait d'entraîner le blocage du processus.

En revanche, il y a un niveau de régulation qui semble plus déterminé et déterministe : la régulation temporelle de l'activation des origines au cours de la phase S. En effet, comme nous le verrons, elle est très robuste, peu flexible pour une situation donnée, et participe à la définition de l'identité-même de la cellule. Cette régulation est appelée « *timing* » de réplication. C'est elle que je vais détailler dans la deuxième partie de cette introduction.

CHAPITRE II : LE *TIMING* DE RÉPLICATION

Les origines de réplication ne sont pas toutes activées au même moment au cours de la phase S, et l'ordre d'activation, ou *timing* de réplication, est un programme robuste qui fait partie de l'identité de la cellule. Ainsi, des origines sont activées en début de phase S, d'autres en milieu de phase S et d'autres en fin de phase S et, définissant ainsi des domaines de réplication précoces, intermédiaires, ou tardifs, respectivement (figure 5).

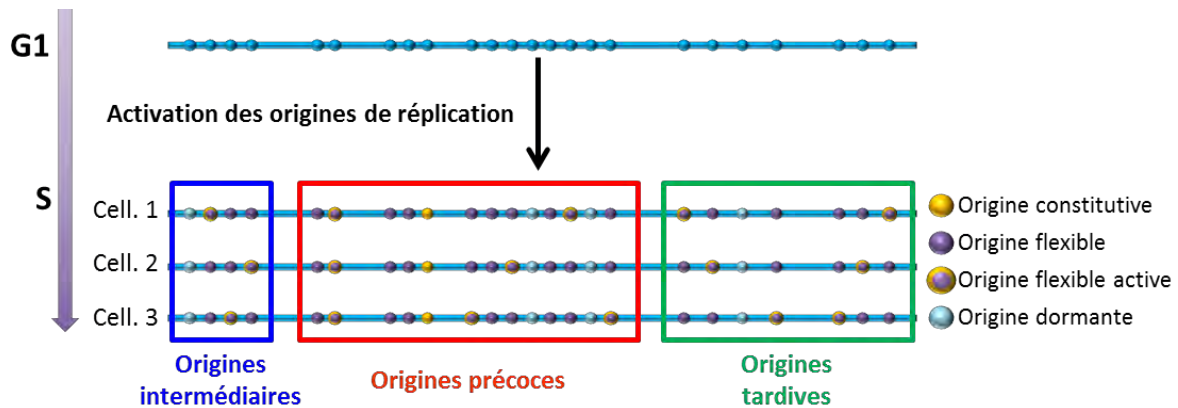


Figure 5 : Régulation temporelle des origines de réplication. Les origines de réplication ne sont pas toutes activées en même temps au cours de la phase S. Elles sont regroupées en domaines synchrones qui peuvent être précoces (cadre rouge), tardifs (cadre vert), ou intermédiaires (cadre bleu), selon que les origines sont activées en début, fin ou milieu de phase S.

I. L'ANALYSE DU *TIMING* DE RÉPLICATION

A. Les foyers de réplication

Très tôt, l'analyse des foyers de réplication par immunofluorescence a suggéré que la réplication est organisée différemment en fonction du stade de la phase S (Figure 6). On distingue ainsi trois profils différents. En début de phase S, la réplication s'organise au sein de multiples foyers de petite taille, répartis un peu partout dans le noyau. En milieu de phase S, ces foyers ont tendance à se relocaliser au niveau de la membrane nucléaire ou à la périphérie des nucléoles. Enfin, en fin de phase S, on observe de plus gros foyers qui correspondent à un regroupement de foyers observés en début et milieu de phase S. Si cela est moins tranché pour les foyers de début et milieu de phase S, il semble que les foyers de fin de phase S se situent dans zones d'hétérochromatine (Nakayasu and Berezney, 1989).

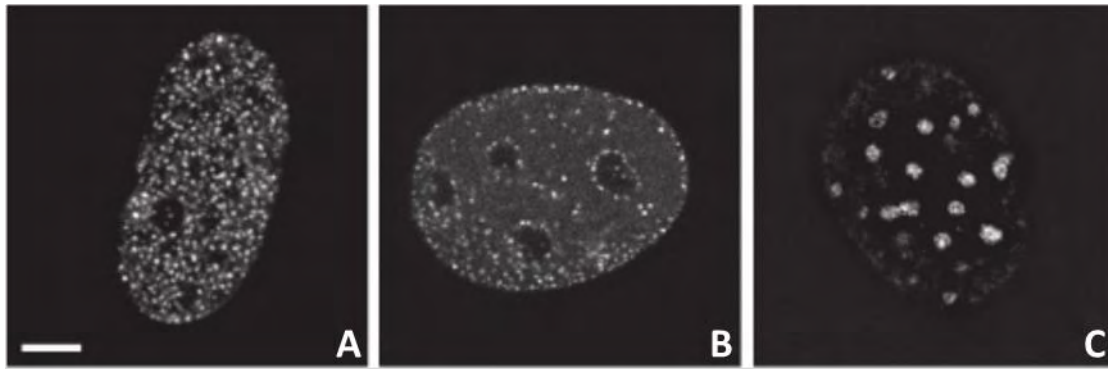


Figure 6 : Distribution des foyers de réplication au cours de la phase S. Image tirée de (Baddeley et al., 2010) . Les foyers de réplication ont été visualisés dans les cellules murines C2C12, exprimant de façon stable PCNA fusionné à la GFP. On distingue trois profils différents : (A) Cellules en début de phase S, (B) Cellules en milieu de phase S, (C) Cellules en fin de phase S.

De plus, le marquage par immunofluorescence de l'ADN synthétisé au même moment de la phase S sur deux générations consécutives, montre que les profils de réplication sont conservés d'une génération à l'autre, et que les mêmes sites sont répliqués au même moment (Ma et al., 1998) . Cette dernière expérience suggère donc que chaque séquence d'ADN est répliquée à un moment spécifique de la phase S, et que le moment d'activation pour chaque origine de réplication n'est pas aléatoire. La structure de ces foyers et leur fonction biologique ne sont pas vraiment établies. Il a été proposé que ces foyers abritent plusieurs réplicons consécutifs, dont l'initiation synchrone dépend d'un seul même complexe initiateur. Mais de récentes études en microscopie à haute résolution semblent indiquer que ces foyers correspondent à un regroupement de plusieurs complexes de réplication, qui seraient activés en même temps (Baddeley et al., 2010; Cseresnyes et al., 2009). C'est également ce que suggère le modèle proposé par Berezney en 2000 et repris par Cayrou *et al* en 2011, selon lequel les réplicons sont regroupés en *cluster*, au sein desquels ils sont répliqués de manière synchrone (Figure 4B).

B. L'étude du *timing* de réplication à l'échelle du génome entier

Si les profils de réplication en immunofluorescence démontrent l'existence de cette régulation temporelle des origines de réplication, c'est le développement de son analyse à l'échelle génomique qui a permis de mieux le comprendre.

1. Méthodes d'étude

Le programme du *timing* de réplication consiste à fixer le moment de duplication des différents domaines chromosomiques du génome. Deux principales méthodes existent pour étudier ce programme. La première (Figure 7A) conçue par l'équipe de David Gilbert

(Gilbert, 1986; Hiratani et al., 2008), repose sur la purification et l'identification de l'ADN néosynthétisé à différents moments de la phase S. Une première étape consiste à révéler l'ADN en cours de synthèse pendant la phase S grâce à l'incorporation de BromodéoxyUridine (BrdU). La technique nécessite aussi de séparer les cellules en fonction de leur position dans le cycle cellulaire par tri en cytométrie en flux (ou FACS, pour « *Flow Analysis Cell Sorting* »). En effet, dans les cellules de métazoaires, les techniques de synchronisation *in vivo* sont souvent lourdes, spécifiques à chaque type cellulaire et font le plus souvent appel à des drogues qui perturbent la réplication, et qui donc vont potentiellement perturber le programme temporel de réplication. Au contraire, le tri par FACS à partir du contenu en ADN est facile à réaliser et reproductible. Après fixation des cellules, l'étape suivante consiste à trier les cellules fixées en fonction leur progression dans le cycle cellulaire, déterminée par incorporation d'Iodure de Propidium (PI). On récolte ainsi séparément les cellules en début et des cellules en fin de phase S. Sur chaque fraction, après sonication, on isole l'ADN qui était en cours de synthèse au moment de la récolte par une immuno-précipitation (IP) dirigée contre le BrdU. Le matériel récolté est amplifié, marqué par des fluorochromes différents selon la fraction dans lequel il se trouvait et hybridé sur puce, qui peut couvrir un chromosome d'intérêt ou bien tout le génome. L'analyse bioinformatique de l'intensité de chaque fluorochrome obtenu pour une région du génome donnée permet de déterminer le moment auquel elle a été répliquée dans la phase S.

Une variante de cette méthode consiste à isoler plus de deux fractions de la phase S, permettant de discriminer avec plus de précision le moment de réplication d'une région d'intérêt au cours de la phase S. Il n'y a alors pas de couplage de l'ADN immuno-précipité avec des sondes, mais une amplification par PCR (« *Polymerase Chain Reaction* ») de la région du génome étudiée dans chaque fraction pour déterminer dans laquelle elle était représentée sous forme d'ADN néosynthétisé (exemple (Hassan-Zadeh et al., 2012)).

Une deuxième technique (Figure 7B) repose sur l'augmentation croissante de la quantité d'ADN au cours de la phase S (Woodfine et al., 2004). Ainsi, une région répliquée précocement va être plus représentée dans une population en phase S qu'une région répliquée tardivement. Cette méthode ne nécessite pas d'incorporation de BrdU, ni d'amplification du matériel génétique. Elle consiste en trier les cellules en phase G1 des cellules en phase S pour en isoler l'ADN. Après sonication, cet ADN est marqué différemment en fonction de la phase du cycle dont il est issu, puis hybridé sur puce. Le calcul du ratio d'intensité S/G1 pour un point du génome donné indique son *timing* de réplication : les régions précoces ont un ratio proche de 2, et plus une région est tardive, plus ce ratio tendra vers 1. Il serait probablement

plus rigoureux de rajouter une étape d'incorporation de BrdU dans cette méthode, ce qui permettrait de trier les cellules non seulement en fonction de leur contenu en ADN, mais aussi en fonction de leur marquage BrdU, rendant possible la discrimination des cellules en phase G1 des cellules en début de phase S, ce qui n'est pas le cas actuellement.

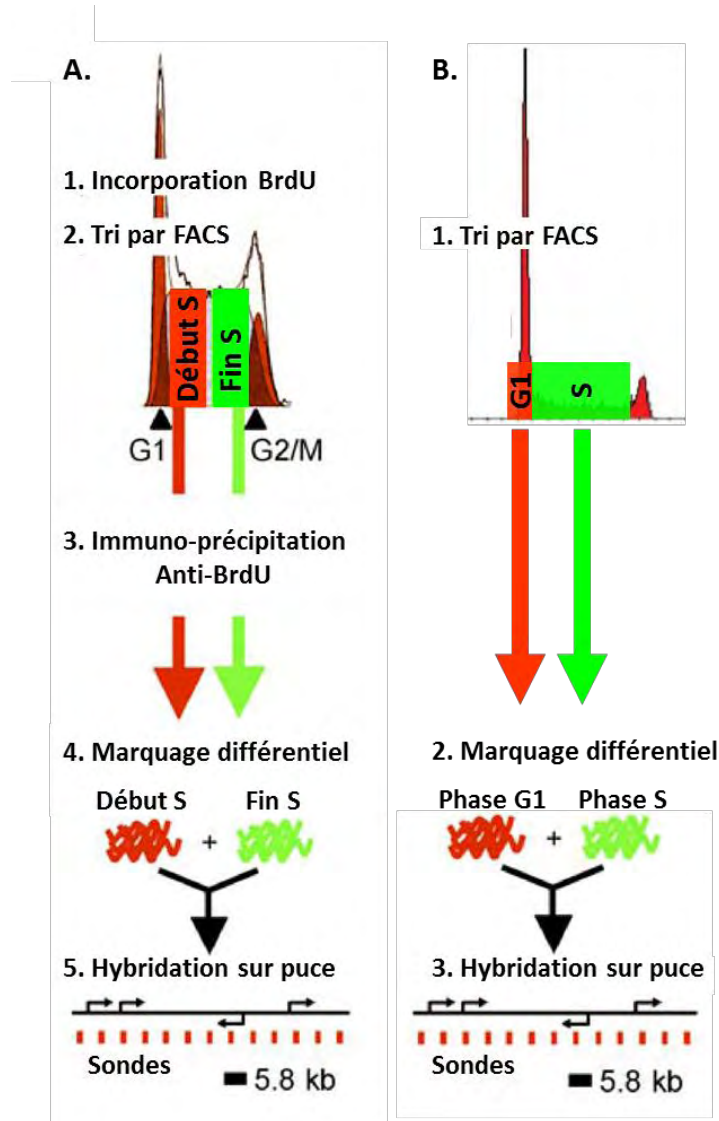


Figure 7 : Méthodes d'étude du *timing* de réplication sur génome entier. (A) Après incorporation BrdU, les cellules de début et fin de phase S sont récoltées. L'ADN néosynthétisé est immuno-précipité avec un anticorps ciblant le BrdU, marqué par des sondes fluorescentes et hybridé sur puce. Adaptation de (Hiratani et al., 2008). (B) Les cellules en phases G1 et S sont récoltées. Dans chaque fraction, l'ADN est isolé et marqué par des sondes fluorescentes, puis hybridées sur puces. Adaptation de (Woodfine et al., 2004).

L'ensemble de ces méthodes et leur protocole associé, de la préparation des cellules à l'analyse bioinformatique, est décrit dans un article publié par l'équipe de Gilbert en 2011 (Ryba et al., 2011a).

2. Visualisation du *timing* de réplication

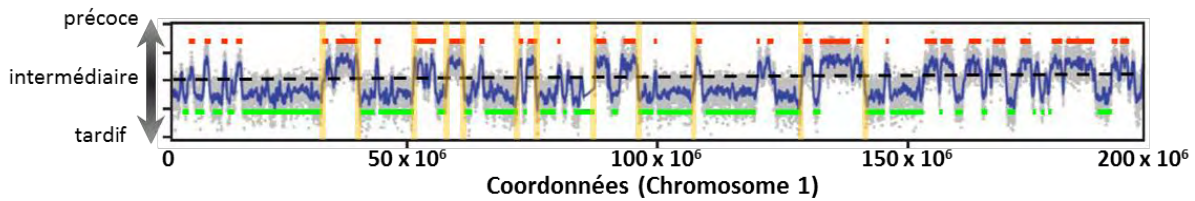


Figure 8 : Profil de *timing* de réplication. Adapté de (Ryba et al., 2011). Le profil de *timing* de réplication donne pour chaque coordonnée du chromosome (ici chromosome 1) indiquée en abscisse, le *timing* de réplication, en ordonnée. On observe ainsi une succession de domaines de réplication répliqués en début (traits rouges), ou en fin (traits verts) de phase S. Ces domaines sont séparés par des régions de transition de *timing*, dont quelques exemples sont montrés en jaune.

Un profil de *timing* de réplication est une représentation graphique du moment de la duplication (en ordonnées) en fonction de la position sur le génome du domaine dupliqué (en abscisses). Il correspond, pour un point donné, au logarithme du rapport entre l'intensité du fluorochrome marquant l'ADN néosynthétisé issu des cellules en début de phase S, et l'intensité de celui correspondant à la fin de phase S. Si le rapport est positif, le domaine est précoce (ou *Early*), s'il est négatif, le domaine est dit tardif (ou *Late*). Si le résultat est proche de 0, la région est intermédiaire (ou *Middle*). Enfin, plus il est élevé, plus le domaine est précoce, plus il est faible, plus le domaine est tardif (Figure 8).

Au final, le profil consiste donc en une succession de grandes régions (de l'ordre du megabase) dont le *timing* est constant (CTR pour « *Constant timing Regions* »), ou domaines de réplication, séparé par des régions de transition (TTR, pour « *Transition Timing Regions* »), dans lesquelles le *timing* bascule. Les CTR correspondent à des segments où l'activation des origines de réplication est synchrone, que ce soit en début ou en fin de phase S. Chez la levure, des travaux suggèrent que la différence entre des origines précoces et des origines tardives viendrait de leur efficacité d'activation, les premières étant plus efficaces que les secondes (Yang et al., 2010). Et de tous récents travaux réalisés par la même équipe viennent de démontrer que comme ils le supposaient déjà, cette efficacité est induite par un nombre plus important d'hélicases MCM recrutées dans le pré-RC au niveau des origines précoces (Das et al., 2015). Cette richesse en MCM permettrait d'augmenter les chances pour ces origines de recruter les facteurs limitants d'activation des pré-RC, tels les DDK, Cdk ou Cdc45. En revanche, les TTR semblent concerner des régions pauvres en origines, répliquées passivement par une fourche unidirectionnelle émanant d'une origine précoce, observée grâce à la technique de molécule unique SMARD (« *Single Molecule Analysis of Replicated DNA* ») (Desprat et al., 2010; Norio et al., 2005). D'autre part, les études du *timing* de réplication sur

le génome entier montrent une corrélation linéaire entre la distance couverte par ces régions et le temps nécessaire à leur réplication, en accord avec l'hypothèse d'une seule fourche unidirectionnelle responsable de la réplication de ces régions de transition (Hiratani et al., 2008; Ryba et al., 2010). En outre, une origine normalement active insérée au sein d'un TTR perd son activité. Et l'induction de la transcription ou de modifications épigénétiques favorables à l'activité des origines n'ont pas permis d'induire l'activation des origines dormantes du TTR (Guan et al., 2009).

Ce modèle a été cependant contredit par une étude de peignage moléculaire (Guilbaud et al., 2012) montrant que la vitesse de réplication dans la majorité des TTR est supérieure à la vitesse maximale de progression d'une seule fourche de réplication. Cette étude révèle la présence de fourches bidirectionnelles asymétriques, suggérant une activation en cascade des origines de réplication au sein des TTR, corrélée à une modification de l'environnement chromatinien, allant d'un état ouvert de la chromatine à l'extrémité précoce du TTR, à un état fermé à son extrémité tardive.

Les profils en U qui correspondent à deux TTR qui se rejoignent, sont la signature de régions pauvres en origines de réplication, dont certaines correspondent à des sites fragiles, dont le cœur de séquence est dépourvu en origines, même dormantes (Letessier et al., 2011; Le Tallec et al., 2012).

II. LES CARACTÉRISTIQUES GÉNÉTIQUES ET ÉPIGÉNÉTIQUES LIÉES AU TIMING DE RÉPLICATION

L'ensemble des études de *timing* de réplication à l'échelle génomique ont permis d'apporter un éclairage sur les différents paramètres influençant le programme temporel d'activation des origines de réplication.

A. La séquence d'ADN

Chez les mammifères, le génome est organisé en isochores, grandes régions du génome regroupées en cinq classes, en fonction de leur composition en GC (Bernardi, 2000). De plus, il existe aussi une corrélation entre la classe de l'isochore et le contenu en gène, les isochores les plus riches en GC étant aussi les plus riches en gènes (Costantini et al., 2009). Enfin, les isochores riches en GC sont pauvres en séquence répétée LINE1 (pour « *Long Interspersed Nuclear Element 1* »), qui sont de longues séquences de plusieurs kb correspondant à des rétrotransposons (Babushok and Kazazian, 2007). De façon intéressante, il existe une corrélation entre isochore et *timing* de réplication (Costantini and Bernardi, 2008). Ces

résultats sont corroborés par plusieurs études sur génome entier qui ont montré un lien entre la séquence primaire d'ADN et le *timing* de réplication. En particulier, les régions répliquées en début de phase S sont riches en GC, enrichies en gènes et pauvres en séquences répétées LINE1, et inversement pour les régions tardives (Hiratani et al., 2008; Ryba et al., 2010). Et de façon intéressante, les régions propices à un changement de *timing* au cours d'un processus biologique donné présentent des niveaux intermédiaires pour ces trois caractéristiques (figure 9).

	Génome des Mammifères			
Isochore: GC%	Haut	Moyen	Faible	Faible
Densité en gènes	Haut	Faible	Moyen	Faible
Densité en LINE1	Faible	Haut	Moyen	Haut
Timing de réplication	Précoce	Tardif vers Précoce	Précoce vers Tardif	Tardif

Figure 9 : Corrélation entre *timing* de réplication et caractéristiques des séquences nucléotidiques.

Les régions précoces sont riches en GC et en gènes, mais pauvre en séquence LINE1, et inversement pour les régions tardives. Les régions propices à changer de *timing* de réplication présentent des niveaux intermédiaires. Adapté de (Rhind and Gilbert, 2013)

Cependant, la séquence par elle-même ne suffit pas à déterminer son *timing* de réplication puisque deux locus homologues peuvent être répliqués à différents moments de la phase S (Farkash-Amar et al., 2008).

B. Relation entre *timing* de réplication et transcription

Les foyers de réplication observés en immunofluorescence suggèrent l'existence d'un lien entre *timing* de réplication et transcription. En effet, comme nous l'avons vu, le profil de fin de phase S correspond à des foyers de réplication situés dans des zones d'hétérochromatine, dans laquelle la chromatine est hyper-condensée, et par conséquent l'activité transcriptionnelle très faible. De plus, nous venons de voir qu'il existe une corrélation entre densité de gènes et *timing* de réplication. Cependant, à l'instar des foyers de réplication de début de phase S qui se situent à la fois dans des zones d'euchromatine et d'hétérochromatine, le lien entre *timing* de réplication et activité transcriptionnelle est plus complexe. Globalement, les gènes actifs sont situés dans les régions précoces et une forte corrélation avec la transcription a été observée dans de nombreuses études (Hiratani et al., 2008; MacAlpine et al., 2005; Woodfine et al., 2004). Cependant des gènes actifs sont aussi

retrouvés dans les régions tardives et inversement il y a dans les régions précoces des gènes inactifs (Farkash-Amar et al., 2008; Hiratani et al., 2008). En fait, la corrélation entre début de phase S et transcription abondante n'est vraie qu'à partir du dernier tiers de la phase précoce de la phase S (Figure 10).

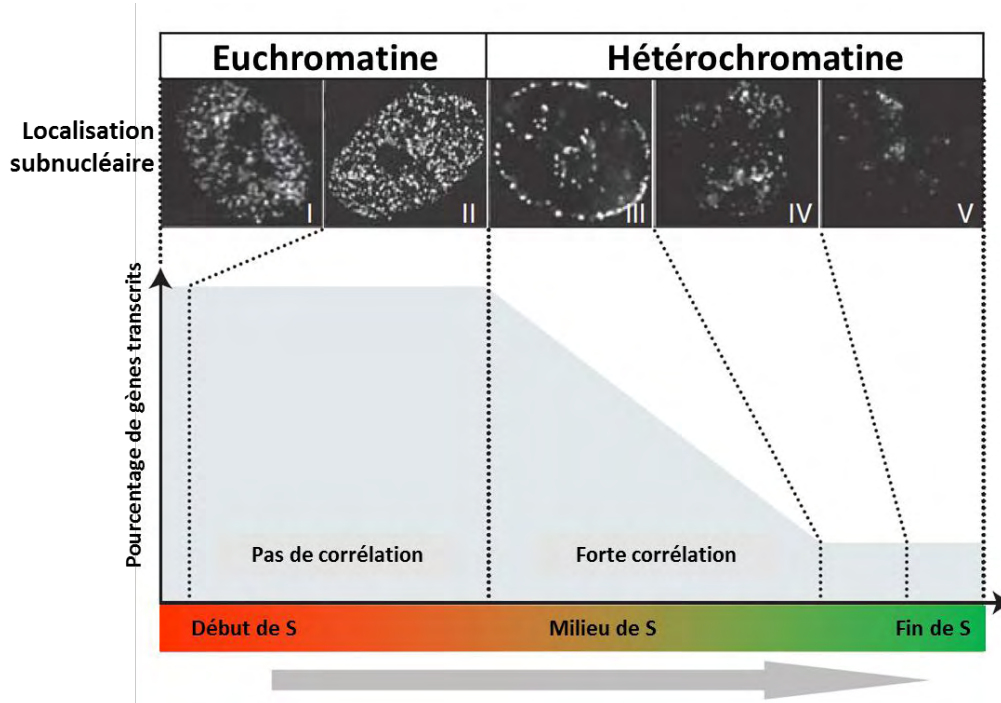


Figure 10 : Lien entre transcription et réplication *timing*. A partir du milieu de la phase S, il existe une corrélation entre transcription et *timing* de réplication, l'activité transcriptionnelle diminuant au fur et à mesure de la phase S. Adapté de (Rhind and Gilbert, 2013)

Une étude dans les cellules de poulet illustre cette relation complexe entre transcription et *timing* de réplication (Hassan-Zadeh et al., 2012). En effet, elle montre qu'une origine flanquée du site de liaison du facteur de transcription USF1 (pour « *Upstream Stimulatory Factor* ») contenu dans l'« *insulator* » HS4 et introduite dans un domaine de réplication tardif voit sa réplication avancée. Cette modification du *timing* de réplication ne nécessite pas qu'une activité transcriptionnelle y soit associée bien qu'elle soit amplifiée lorsqu'un gène est transcrit à proximité. De façon intéressante, l'insertion de cette séquence induit une modification de l'environnement chromatinien : les marques répressives H3K27me3 majoritaires avant l'insertion de l'*insulator* disparaissent au profit de marques de chromatine active (H3K4me2, H3 et H2A.Z acétylés) suite au recrutement d'USF1 après insertion. Cette étude permet donc non seulement l'identification d'un nouvel élément *cis* participant à la régulation temporelle des origines de réplication, mais aussi illustre le lien qui existe entre épigénétique et *timing* de réplication.

C. Environnement chromatinien

1. A l'échelle du nucléosome

Le lien entre *timing* de réplication et transcription suggère qu'il existe aussi un lien avec l'environnement chromatinien et les marques d'histones. Et en effet, le *timing* de réplication est fortement associé à l'accessibilité de la chromatine (Bell et al., 2010), caractéristique des domaines précoces. Une étude récente montre que les profils d'hypersensibilité à la DNase sont les prédictors les plus forts du *timing* de réplication dans les cellules somatiques (Gindin et al., 2014). De plus, dans plusieurs études, des marques de chromatines actives ont été associées à ces domaines répliqués en début de phase S : H3K4me3, H2AZ, H3K9ac H3K36me3 ainsi que H4K20me1 (Bell et al., 2010; Picard et al., 2014). Chez la drosophile, l'acétylation de l'histone H4, H4K16ac est aussi associée aux origines précoces (Comoglio et al., 2015). Dans les cellules humaines K562, les régions répliquées en milieu de phase S sont associées à un duo d'histones méthylées : H4K20me1 and H3K27me3 (Picard et al., 2014). Enfin, la marque répressive qui semble la plus caractéristique des domaines de réplication de fin de phase S est H3K9me3 (Hiratani et al., 2008; Picard et al., 2014; Ryba et al., 2010). Les régions très tardives sont aussi associées à la protéine HP1, qui induit la formation d'hétérochromatine. Cependant, les coefficients de corrélation ne sont pas toujours très bons et certaines marques sont retrouvées dans différentes phase de la phase S. Enfin, il existe de nombreuses origines qui ne présentent pas de marque d'histone spécifique, en particulier dans des régions réprimées, répliquées dans la deuxième moitié de la phase S. (Julienne et al., 2013; Picard et al., 2014). Le *timing* de réplication n'est donc pas directement ou uniquement corrélé à ces modifications d'histones.

2. A l'échelle du noyau, organisation spatiale

Les différents profils de réplication observés en IF (figures 5 et 9) suggèrent une organisation spatiale des domaines de réplication en fonction de leur *timing* de réplication. Les régions précoces sont au cœur du noyau, alors que les régions tardives se situent à la périphérie du noyau ou des nucléoles. Cette localisation des régions tardives à la périphérie semble résulter d'un ancrage à la matrice nucléaire par l'intermédiaire des lamines (Guelen et al., 2008; Hansen et al., 2010) (Figure 11B).

En 2009, une nouvelle méthode est mise au point : le Hi-C, qui correspond au couplage de la méthode des 3C (pour « Chromosome Conformation Capture ») avec le séquençage des produits de ligation obtenus. Cette méthode a permis d'identifier les interactions présentes

dans la chromatine à l'échelle du génome entier et ainsi de reconstituer l'organisation en 3 dimensions du noyau. En particulier, cette étude a mis en évidence une organisation du génome en deux compartiments distincts, A et B, dans le noyau. Les séquences contenues au sein d'un compartiment interagissent entre elles, alors qu'il n'y a aucune interaction entre des séquences issues de deux compartiments différents. Ces deux compartiments diffèrent en termes d'ouverture de la chromatine, le compartiment A correspondant à un état ouvert de la chromatine, et le compartiment B à un état fermé. Il existe une forte corrélation ($R=0.8$) entre cette compartimentation du génome et les profils de *timing* de réplication (Ryba et al., 2010), alors que la corrélation avec les marques d'histones est moins marquée (Figure 11A-B). Le *timing* de réplication est donc fortement lié à l'organisation tridimensionnelle de la chromatine au sein du noyau, et les modifications d'histones seraient secondaires dans la détermination du programme temporel de l'activation des origines de réplication.

Ainsi, les régions précoces et régions tardives sont séparées en deux compartiments nucléaires distincts. Des régions chromatinienne regroupées dans un même espace sont répliquées en même temps, et les régions de transition du *timing* de réplication correspondent à un mouvement vers un nouveau domaine chromatinien indépendant.

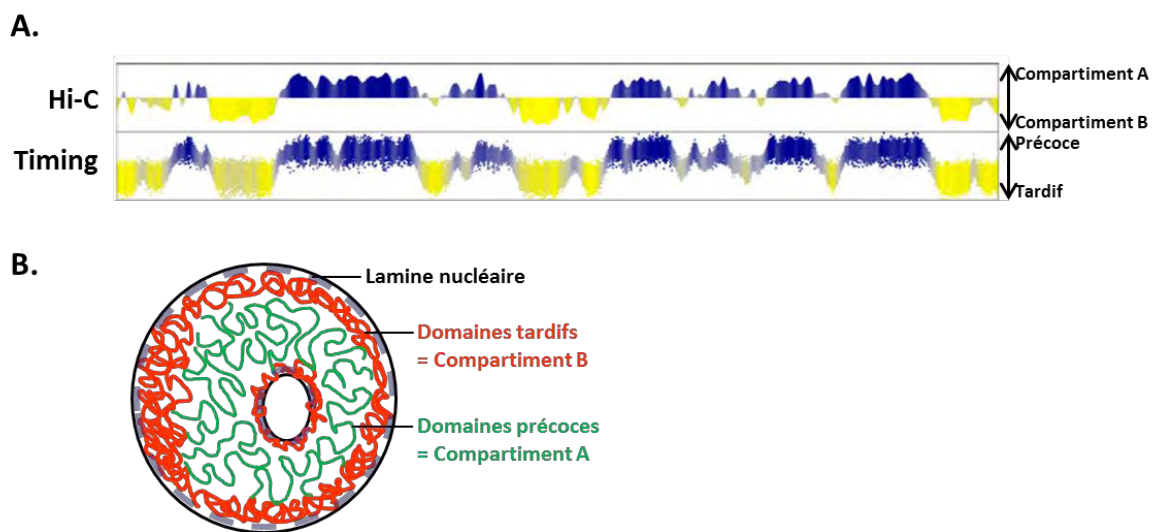


Figure 11 : Organisation spatiale des domaines de réplication. (A) Comparaison entre *timing* de réplication et diagramme de *Eigenvector*, sur lequel apparaît l'enrichissement en compartiment A (bleu) ou B (jaune) pour une région donnée (abscisse) après Hi-C, dans une même lignée (Ryba et al., 2010). (B) Représentation graphique de l'organisation nucléaire des domaines de réplication.

III. MÉCANISMES MOLÉCULAIRES DU *TIMING* DE RÉPLICATION

A. Le point de décision du *timing* de réplication

Le point de décision du *timing* de réplication, ou « *Timing Decision Point* » (TDP) a été mis en évidence en 1999 par l'équipe de D. Gilbert (Dimitrova and Gilbert, 2000). Pour cela, les auteurs ont marqué différemment l'ADN répliqué en début de phase S de celui répliqué en fin de phase S dans des cellules de hamster CHO. Puis ils ont isolé ces noyaux à différents temps dans la phase G1 suivante pour les introduire dans un système *in vitro* compétent pour la réplication, les extraits d'œuf de xénope. Ils ont ensuite déterminé si le *timing* de réplication était conservé par rapport à celui du cycle précédent. Les résultats ont montré que le *timing* était retrouvé systématiquement si l'isolation des noyaux survenait au minimum, 1h à 2h après le début de la phase G1. Si elle intervenait avant, les domaines précoces et tardifs ne sont pas correctement redistribués dans le noyau et le *timing* de réplication est aléatoire. Ils ont ainsi mis en évidence un moment précis dans la phase G1 où le *timing* de réplication est établi, le TDP, qui correspond à la redistribution spatiale des domaines dans le noyau lorsque celui-ci retrouve sa forme interphasique en sortie de mitose. Enfin, ce TDP intervient après le recrutement du dernier composant du pré-RC, les Mcm, mais précède l'ODP.

Quelques années plus tard, l'équipe de Gilbert a utilisé la même approche expérimentale en introduisant des noyaux isolés en phase G2 dans les extraits de xénope compétents pour la réplication. De nouveau, alors que l'organisation spatiale de la chromatine dans le noyau est la même que celle de la phase S, la réplication est temporellement désorganisée. A l'inverse, si la même méthode est appliquée sur des cellules en quiescence, le *timing* de réplication est conservé, alors que l'organisation de la chromatine est altérée par rapport aux cellules cyclantes (Lu et al., 2010). Cela suggère donc que l'organisation de la chromatine ne suffit pas à elle toute seule pour dicter le *timing* de réplication et qu'un ou plusieurs des déterminants, acquis au TDP, sont perdus au cours de la phase S.

B. Mcm et compétition pour les facteurs limitant de l'activation des origines de réplication

Plusieurs études ont montré, chez la levure ou les métazoaires que l'efficacité d'une origine de réplication, et donc sa probabilité d'activation, dépendent de la compétition inter-origines pour les facteurs limitants de l'activation des origines, dont Cdc45 ou les Cdk et Ddk. En effet, chez la levure, la surexpression d'un de ces facteurs permet non seulement

d'augmenter globalement ou localement l'efficacité des origines de réplication, mais aussi perturbe le *timing* de réplication en induisant précocement l'activité d'origines tardives (Mantiero et al., 2011; Patel et al., 2008; Tanaka et al., 2012). Il est intéressant de noter que ces perturbations dans la régulation de l'activité des origines est une source de stress répliatif et d'instabilité génétique. De la même façon, la dérégulation des Cdk chez les eucaryotes supérieurs module l'efficacité des origines de réplication ainsi que le *timing* de réplication (Katsuno et al., 2009). Se pose alors la question de la manière dont se fait le choix des origines actives par les facteurs limitants de l'initiation, ou en d'autres termes, sur quel(s) critère(s) repose la compétition entre les pré-RC mis en place en G1.

Comme évoqué précédemment, les récents travaux de l'équipe de Rhind (Das et al., 2015) apportent certainement un éclairage sur ce point. En effet, ils montrent que le *timing* de réplication est déterminé en partie par le nombre de Mcm recrutés sur une origine : plus l'origine est riche en double-hexamères de Mcm, plus il y a de chances qu'au niveau de cette origine, un d'entre eux soit activé, c'est-à-dire soit la cible d'un de ces facteurs limitants de l'initiation de la réplication, et avec lui l'origine de réplication. Ces origines riches en Mcm seraient donc plus efficaces, et activées en début de phase S (Figure 12). Ce modèle est aussi compatible avec le TDP, puisque celui-ci a lieu à priori après le recrutement de Mcm (Dimitrova and Gilbert, 2000). Et la distribution des Mcm recrutés par origine constituerait alors un mécanisme du TDP. Cependant, on ne connaît toujours pas la régulation du chargement des Mcm. Chez la levure, plus le complexe Orc est recruté tôt à la chromatine en G1, voire en G2 du cycle précédent, plus il est stable et correspondra à une origine précoce ou efficace (Belsky et al., 2015; Wu and Nurse, 2009). On peut imaginer que dans de tels cas, le recrutement des Mcm à leur niveau est aussi plus efficace du fait d'un temps de résidence de Orc plus long à la chromatine.

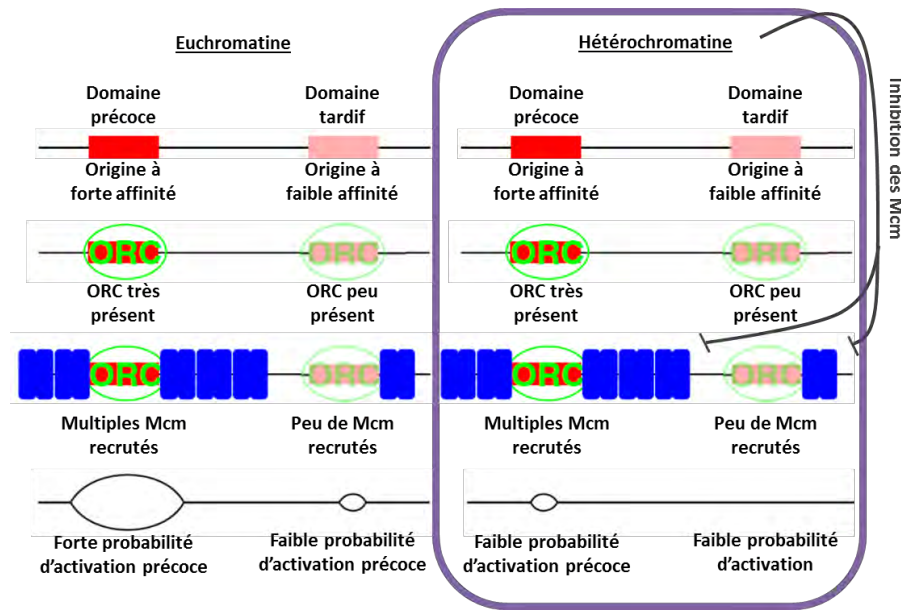


Figure 12 : Modèle de régulation du *timing* de réplication par le nombre de Mcm recrutés au pré-RC.

Adapté de (Das et al., 2015). Lors de la mise en place des pré-RC, les origines sur lesquelles Orc vient se fixer le plus tôt recruteront un plus grand nombre de double-hexamères de Mcm, augmentant la probabilité d'une activation précoce dans un contexte euchromatinien. Dans un contexte hétérochromatinien, un niveau en amont de régulation du *timing* de réplication empêche l'activation des Mcm en début de phase S.

C. Rif1 un régulateur conservé du *timing* de réplication

1. Rif1 régule le *timing* de réplication de la levure à l'homme, en passant par la souris

La première étude à avoir mis en évidence un rôle de Rif1 dans le *timing* de réplication a été réalisée chez la levure (Hayano et al., 2012). Cette découverte est issue de la recherche d'une mutation qui permettrait d'améliorer la survie d'une souche déficiente pour Hsk1, homologue de cdc7 chez la levure et donc membre des DDK, incapable d'activer les origines de réplication. Les auteurs ont donc observé que la délétion de Rif1, connu jusqu'alors pour son rôle dans la stabilité des télomères, permettait de restaurer la viabilité de la souche Hsk1Δ ainsi que la réplication de l'ADN. Leur étude a montré que non seulement Rif1 peut induire l'activation ou l'inhibition d'origines de réplication, mais surtout que sa déplétion entraîne une modification globale du *timing* de réplication par activation précoce d'origines tardives, ou inversement retard dans l'activation d'origines normalement précoces. Ils ont également observé que Rif1 se lie, en début des G1, aux origines ou à proximité de plusieurs origines, y compris sur des régions non-téломériques. Cependant, des expériences de ChIP-chip suggèrent que Rif1 n'influence pas la mise en place du pré-RC, puisque les sites de liaison à MCM ne semblent pas affectés par le statut de Rif1. En revanche, en absence de Rif1, le recrutement de cdc45 est augmenté. L'hypothèse des auteurs est donc que Rif1 régule l'activation des origines de réplication via le recrutement de cdc45. Son rôle dans le *timing* de

réplication passerait par la régulation de l'organisation chromatinienne du noyau au moment du TDP.

Le rôle de Rif1 dans le *timing* de réplication chez les mammifères a été décrit juste après, chez la souris et chez l'Homme (Cornacchia et al., 2012; Yamazaki et al., 2012). Chez la souris (Cornacchia et al., 2012), les auteurs ont d'abord observé en immunofluorescence que Rif1, en milieu de phase S, est organisé en foyers qui ne co-localisent pas avec les fourches de réplication, mais plutôt les précèdent, comme si Rif1 prédisait les foyers de réplication. Ce côté « prédictif » a poussé les auteurs à regarder si Rif1 pouvait intervenir dans le *timing* de réplication. Et effectivement, la déplétion de Rif1 entraîne, à l'échelle du génome entier, de profonds changements de *timing* de réplication dans les deux sens : de précoce à tardif, et inversement. De façon intéressante, cela entraînait dans les fibroblastes une redéfinition des domaines de réplication pour des domaines plus petits, à l'instar de ce qui est observé dans des cellules peu différenciées (Hiratani et al., 2008). Il semblerait donc que, chez la souris, la déplétion de Rif1 abolisse un programme temporel acquis pendant la différenciation. Chez l'homme (Yamazaki et al., 2012), la déplétion de Rif1 induit aussi une perturbation du *timing* de réplication : perte des profils de foyers de réplication de type « milieu de phase S » et modification dans les deux sens des profils de *timing* réalisés sur le chromosome 5. En revanche, contrairement à la souris, la déplétion de Rif1 induit une consolidation, et non une fragmentation des domaines de réplication. Par ailleurs, les auteurs montrent qu'en absence de Rif1, il y a une augmentation de la phosphorylation des Mcm par Cdc7 en entrée de phase S, sans que l'activité de cdc7 soit pour autant affectée. Cette observation est intéressante car chez la levure, Rif1 induit un défaut de phosphorylation des Mcm par l'intermédiaire d'une phosphatase, la protéine Phosphatase 1 (PP1). Cette dernière interagit avec Rif1 et vient déphosphoryler les Mcm phosphorylés par Cdc7 sur les pré-RC ciblés par Rif1 (Davé et al., 2014; Hiraga et al., 2014). Ainsi, Rif1 participerait au TDP en contrecarrant l'activation d'origines intermédiaires par les Ddk en début de phase S (Figure 13A).

L'étude de Yamazaki *et al.* a aussi identifié Rif1 comme un régulateur de l'architecture nucléaire et de l'organisation chromatinienne. Il est recruté à la matrice nucléaire à la transition entre la mitose et la phase G1, co-localise avec les lamines et régule la mise en place des boucles de chromatine. Sa déplétion entraîne une augmentation de la taille des boucles de chromatine, en accord avec une augmentation de la distance inter-origine et une accélération de la progression des fourches en début, et surtout milieu de phase S.

Ainsi, ces différentes études ont identifié Rif1 comme un acteur important dans la régulation spatiale et temporelle des origines de réplication chez les eucaryotes.

2. Recrutement de Rif1 à la chromatine au niveau de structures en G-quadruplex.

L'équipe de H. Masai vient d'identifier un mécanisme de recrutement de Rif1 à la chromatine dans les parties non télomériques chez la levure (Kano et al., 2015). En effet, s'il avait été établi depuis plus de 10 ans que le recrutement de Rif1 aux télomères chez la levure *S. pombe* était dépendant d'un autre facteur, Taz1 (protéine shelterin de liaison aux télomères, homologue de Trf1 humain) (Kano et Ishikawa, 2001), on ne connaissait pas la façon dont il est recruté sur les séquences non télomériques. Dans leur étude, les auteurs montrent que le recrutement de Rif1 dans ces régions-là est indépendant de Taz1, puisque les résultats obtenus après immunoprécipitation de chromatine par ciblage de Rif1 sont les mêmes en présence ou en absence de Taz1. En outre, ces expériences ont permis d'identifier les sites de liaison de Rif1 à la chromatine. Pour la majorité d'entre elles (54 sur 90), ces régions présentent au moins une séquence consensus de liaison riche en G (Figure 13B). Mais c'est pour celles qui en présentent au moins deux, séparées d'une centaine de paires de bases, que la liaison est la plus forte. En outre, ces séquences consensus peuvent *in vitro* former des structures secondaires en G-quadruplex avec de petites séquences riches en G situées à proximité. Il semblerait donc que cette structuration soit importante pour la liaison de Rif1, qui *in vitro*, se lie davantage à des substrats capables de former des G4. Cependant, les sites de liaison de Rif1 ne correspondent pas à la séquence d'ADN standard permettant de former des G4 ($G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$) et une large majorité d'entre eux ne correspondent pas aux sites prédits pour former des G4 chez la levure. C'est d'ailleurs pour cela que la présence de plusieurs séquences consensus et des petites séquences riches en G améliore la liaison de Rif1 à la chromatine. En effet, on peut alors imaginer qu'ils interagissent entre eux pour former les structures secondaires.

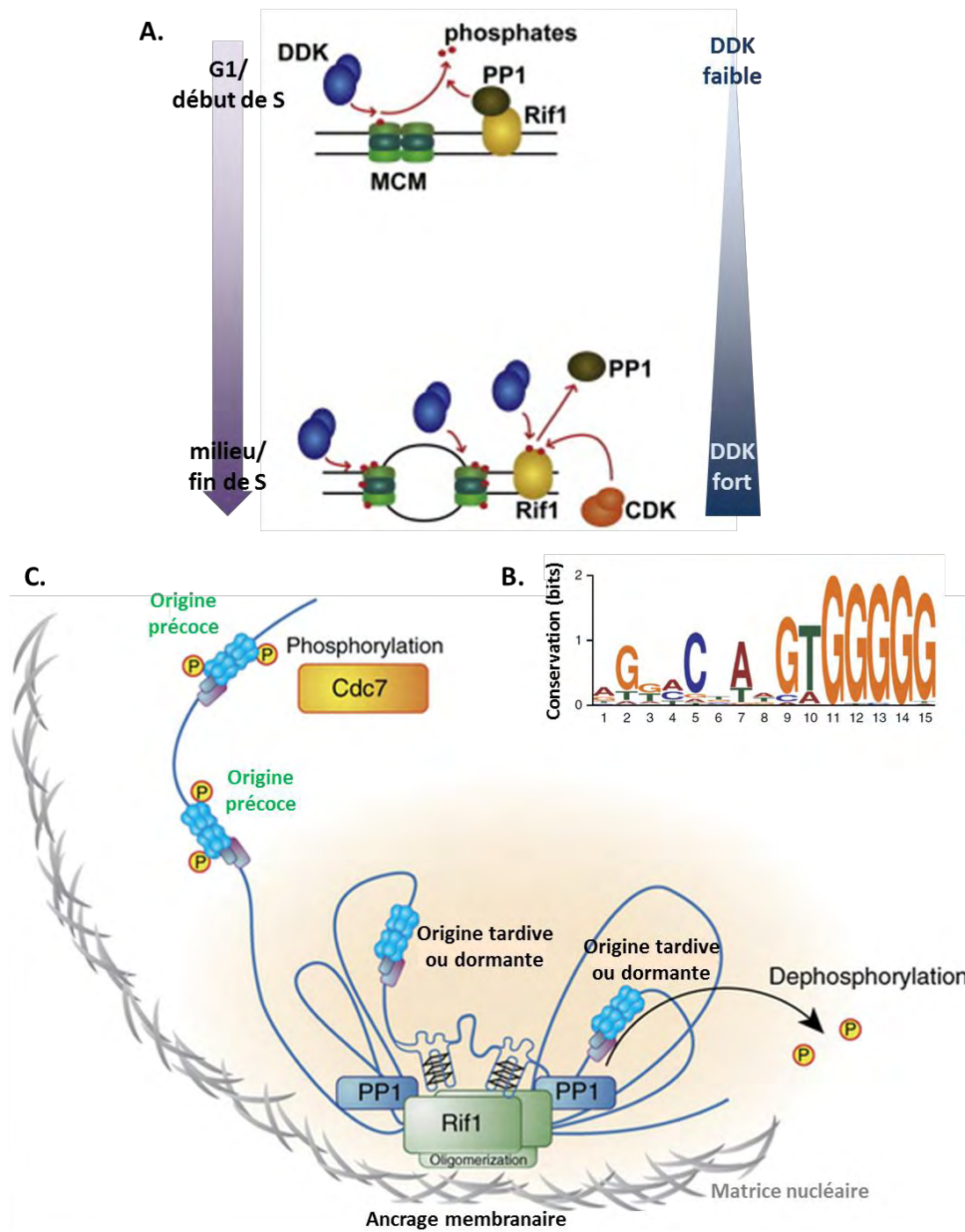


Figure 13 : Modèle pour le mécanisme d'action de Rif1 dans le *timing* de réplication. (A) Régulation de l'inhibition des origines de réplication par Rif1 en fonction de la disponibilité en kinases DDK. Adapté de (Davé et al., 2014). En G1, début de phase S, Rif1 induit la déphosphorylation des Mcm en recrutant PP1 aux origines de réplication. L'origine est maintenue inactive. En milieu et fin de phase S, les DDK sont suffisamment abondantes et disponibles pour phosphoryler les Mcm, mais aussi Rif1, induisant le relargage de PP1. L'origine est activée. (B) Séquence consensus de liaison de Rif1 à la chromatine. (C) Modèle de recrutement de Rif1 au niveau des G4 permettant la formation de domaines de réplication tardifs chez la levure. Tiré de (Kanoh et al., 2015)

Il serait néanmoins intéressant de vérifier *in vivo* si Rif1 se lie bien à des structures secondaires au niveau de ces sites de liaison. Enfin, les auteurs ont montré que ces sites de liaison sont situés à proximité d'origines tardives ou dormantes et que la présence de Rif1 à leur niveau induit une répression de la réplication sur plusieurs dizaines de kilobases. L'ensemble de ces travaux leur ont permis de proposer un modèle pour le mécanisme d'action de Rif1 dans sa régulation du *timing* de réplication au niveau des régions tardives (Figure 13C).

Enfin, il est intéressant de noter, que bien que les auteurs de cette étude ont montré que la liaison de Rif1 dans les bras des chromosomes est indépendante de Taz1, une étude a montré que Taz1 se liait aussi aux parties internes des chromosomes, inhibant la réplication sur quelques dizaines de kb et régulant ainsi leur *timing* de réplication en maintenant ses régions tardives. Cette inhibition de la réplication est due à l'inhibition de la liaison de Sld3, homologue de la Treslin humaine, au niveau des origines empêchant ainsi leur activation. Enfin, les régions dont la réplication est activée en absence de Taz1 correspondent à celles activées en absence de Rif1 (Tazumi et al., 2012). L'ensemble de ces résultats suggèrent donc qu'il peut exister une catégorie d'origines reconnue par Rif1 et Taz1, qui coopèrent pour maintenir un *timing* retardé, et dont le recrutement de Rif1 serait indépendant des G4.

IV. IMPORTANCE BIOLOGIQUE DU TIMING DE REPLICATION

Le *timing* de réplication est une marque épigénétique à part entière (Hiratani and Gilbert, 2009). Sa conservation au cours de l'évolution (Ryba et al., 2010) suggère qu'il joue un rôle biologique important.

A. Le *timing* de réplication, « carte d'identité de la cellule »

1. Transmission d'une génération à l'autre

Le *timing* de réplication est un programme cellulaire très robuste. En effet, pour un type cellulaire donné la position et le moment de réplication des foyers de réplication, analysés par de courtes incorporations d'analogues de bases, est conservé au cours des divisions cellulaires (Dimitrova and Gilbert, 2000; Ma et al., 1998; Sadoni et al., 2004). L'organisation nucléaire est donc maintenue pendant toute la durée de l'interphase, et transmise aux cellules filles. Au cours du développement, cette distribution spatiale des domaines en fonction de leur *timing* de réplication participe au processus même du développement, comme le montre l'exemple du chromosome X. En effet, l'inactivation d'un des deux chromosomes X chez la femelle est

associée à une désynchronisation programmée de la réplication de chaque chromosome qui sera conservée au cours des générations cellulaires. Celui mis en silence devient alors répliqué tardivement. Ce retard de réplication s'accompagne de modifications épigénétiques importantes, induisant un état d'hétérochromatine facultative (Heard and Distèche, 2006). Une fois qu'un des deux chromosomes a été inactivé, c'est le même qui le restera au fur et à mesure du développement.

2. Le *timing* de réplication, signature de l'état de différenciation et du type cellulaire

L'exemple du chromosome X montre aussi que le *timing* de réplication évolue au cours du développement et de la différenciation. Une étude à grande échelle, sur le génome entier de plusieurs lignées cellulaires de souris au cours d'un processus de différenciation, a montré qu'un profond changement dans le *timing* de réplication accompagne ce processus, environ 20% du génome étant affecté (Hiratani et al., 2008). Dans cette étude, les auteurs ont induit la différenciation en cellules neurales, et ce en utilisant plusieurs protocoles différents, de cellules souches embryonnaires de souris : deux souches issues de lignées différentes, et une induite *in vitro* (iPS). Ils ont observé que le *timing* de réplication signe un état de différenciation. Non seulement le *timing* de réplication est quasiment identique d'une cellule-souche à l'autre malgré des origines très différentes, mais il est aussi très conservé entre deux cellules neurales différenciées ayant subi des protocoles de différenciation différents. Cette modification du programme temporel d'activation des origines se caractérise majoritairement par le regroupement de petits domaines asynchrones en grand domaines répliqués en même temps après différenciation, appelé « consolidation ». Mais il a aussi été observé des modifications dans les domaines de transition du *timing*, ainsi que quelques cas dits d'« isolation » correspondant à l'apparition de petits domaines de réplication isolés. La « consolidation » des domaines est à l'origine de la correspondance entre domaines de réplication et isochores décrite précédemment : les régions précoces sont riches en GC et en gènes, et pauvre en LINE1, et inversement pour les régions tardives. Et les régions dont le *timing* de réplication est modifié au cours de la différenciation présentent des caractéristiques intermédiaires pour ces trois facteurs (Figure 9). De plus, ces changements de *timing* de réplication corrélaient significativement avec l'activité transcriptionnelle. En effet, dans une majorité des cas, les gènes contenus dans des domaines dont le *timing* est avancé vont être activés, alors que ceux situés sur des domaines retardés sont réprimés. Enfin, ces évolutions dans le *timing* de réplication sont associées à une relocalisation nucléaire des domaines affectés : les domaines avancés migrent de la périphérie vers le centre du noyau et

inversement pour les domaines retardés. Ces observations ont également été rapportées chez l'Homme, montrant une conservation du processus d'adaptation du *timing* de réplication pendant la différenciation entre différentes espèces chez les mammifères (Ryba et al., 2010). On peut donc dire que le *timing* de réplication « signe » l'état de différenciation d'une cellule et les travaux de Ryba *et al.* ont montré qu'il permet de discriminer un type cellulaire d'un autre (Ryba et al., 2011b). Dans cette étude, les auteurs ont mis au point une méthode permettant de déterminer le type cellulaire à partir d'une sorte de « *timing* de réplication minimal », qui, à l'instar des « empreintes digitales », surnom attribué à ces domaines de réplication caractéristiques d'un type cellulaire donné, permettent d'identifier un tout à partir d'une partie.

B. *Timing* de réplication et cancer

Nous venons de voir que le *timing* de réplication constitue une marque épigénétique à part entière, marqueur de l'identité cellulaire, soulignant l'importance de ce processus dans la « destinée » d'une cellule. De façon intéressante, plusieurs études ont montré que ce marqueur si robuste était altéré dans plusieurs maladies, maladies génétiques et surtout cancer présentent un *timing* de réplication altéré (Pour revue (Donley and Thayer, 2013)). Dans le syndrome de Roberts ou la maladie de Gilles de la Tourette (Barbosa et al., 2000; Grealley et al., 2003), une désynchronisation entre la réplication des deux allèles de régions d'intérêt, qui sont synchrones chez le sujet sain a été observée. Une étude globale a aussi montré que les SNP, ou de nombreux gènes associés à des maladies telles le cancer ou Alzheimer (par exemple les oncogènes Cycline D1, FGF (« *Fibroblast Growth Factor* ») 3 et 4 ou APP (« *Amyloid Beta (A4) Precursor Protein* ») respectivement), sont situés dans des zones de transition du *timing* de réplication (Watanabe et al., 2002), zones dépourvues ou appauvries en origines de réplication et qui sont souvent le nid de sites fragiles. Mais la maladie pour laquelle le lien avec le *timing* de réplication est le plus décrit est certainement le cancer. En effet, de nombreuses études à la fin des années 1990, début des années 2000 ont aussi décrit une désynchronisation de la réplication de loci normalement synchrones dans les leucémies, mais aussi dans des lésions pré-cancéreuses ou dans les lymphocytes de patients atteints de tumeurs solides (Amiel et al., 1998a, 1999, 1998b; Cytron et al., 2011; Korenstein-Ilan et al., 2002). Dans le cancer du sein, certains gènes impliqués dans la pathologie (TP53, RAD51, ATM, PTEN, et cMYC) subissent un retard de *timing*, associé à une perte de leur expression (Fritz et al., 2013). En 2012, une étude sur le génome entier de plusieurs patients atteints de leucémie aigue lymphoblastique a montré que le *timing* de réplication, conservé entre des

lymphoblastes sains, et altéré sur l'ensemble du génome chez les patients avec une tendance à une avancée du *timing* chez la plupart d'entre eux. Si le *timing* de réplication entre les différents patients présente des différences substantielles, il existe des modifications communes à tous les patients, que les auteurs regroupent sous le nom d'« empreintes digitales pan-leucémiques ». Il est possible que ce soit les témoins d'évènements communs précurseurs dans la maladie, conservés au cours de la tumorigenèse. De façon intéressante, cette étude a aussi montré que de nombreux changements de *timing* associés à la maladie ont lieu près de sites de réarrangement ou translocations chromosomiques caractéristiques de la maladie, suggérant que la modification du *timing* pourrait favoriser la survenue de tels évènements (Ryba et al., 2012).

Le lien entre *timing* de réplication et instabilité génétique a été décrit à de nombreuses reprises. Tout d'abord, dans un contexte normal, les régions répliquées en fin de phase S sont beaucoup plus susceptibles aux mutations ponctuelles ou à la survenue de duplication génique (Donley and Thayer, 2013; Lang and Murray, 2011; Stamatoyannopoulos et al., 2009). De plus, des études suggèrent qu'une dérégulation du *timing* de réplication peut entraîner des problèmes dans la condensation des chromosomes, et par voie de conséquence une mauvaise ségrégation et des translocations en mitose (Breger et al., 2004; Loupart et al., 2000; Smith et al., 2001). Enfin, comme évoqué précédemment, les sites fragiles, dont l'expression constitue un évènement précoce dans la tumorigenèse sont souvent situés au niveau des régions des TTR. Ainsi, le *timing* de réplication et ses altérations peuvent jouer un rôle moteur dans l'instabilité génétique qui accompagne le développement tumoral, en alimentant le phénotype mutateur par mutations ponctuelles ou en favorisant les amplifications, délétions ou translocations chromosomiques.

Ainsi le *timing* de réplication est un programme robuste, marque épigénétique à part entière qui définit l'identité d'une cellule, mais aussi son « état de santé ». Sans que l'on puisse affirmer que son altération soit à l'origine du développement tumoral, ou si c'est la maladie qui induit des modifications, nul doute qu'il participe à la tumorigenèse. Et les résultats de Ryba *et al.* suggèrent que c'est un évènement qui survient à des stades précoces de la maladie.

RÉSULTATS

Objectifs de mon travail de thèse.

Quand j'ai commencé ma thèse, mon projet visait à comprendre quel(s) rôle(s) pouvait jouer une ADN polymérase spécialisée, l'ADN polymérase Thêta (Polθ) dans la réplication génomique. En effet, comme nous le verrons plus tard, mon équipe a montré que cette ADN polymérase présente une signature toute particulière dans les cancers du côlon, du poumon et du sein (Allera-Moreau et al., 2012; Lemée et al., 2010; Pillaire et al., 2010). Nos recherches visaient donc à explorer, comme c'était le cas pour deux autres ADN polymérases spécialisées, les ADN polymérases eta (η) et Kappa (κ), si elle intervenait dans la réplication du génome en absence de stress exogène, et de quelle manière. Au fur et à mesure, les résultats expérimentaux nous ont orientés vers un versant complètement inattendu de la réplication : la régulation temporelle de l'activation des origines de réplication. Par la suite, j'ai donc continué à chercher ce qui pouvait influencer ce *timing* de réplication, programme si robuste, et pièce d'identité de la cellule. En particulier, compte tenu du lien particulier qui existe entre dérégulation du *timing* de réplication et cancer, j'ai cherché à évaluer l'impact que pouvait avoir le stress réplicatif, un des moteurs de la cancérogenèse, et caractéristique intrinsèque de la cellule tumorale sur ce programme temporel des origines de réplication.

CHAPITRE I : DÉCOUVERTE D'UN NOUVEAU RÉGULATEUR DU TIMING DE RÉPLICATION : L'ADN POLYMÉRASE THÊTA

I. Les ADN polymerases

A. Deux grandes familles d'ADN polymérases

Les ADN polymérases, responsables de la duplication de la double hélice de l'ADN, sont regroupées en deux grandes familles. D'une part, les ADN polymérases dites répliquatives, la primase Pol alpha ($Pol\alpha$), Pol delta ($Pol\delta$) et Pol epsilon ($Pol\epsilon$), sont impliquées dans la réplication fidèle du génome au cours de la phase S en absence de lésions externes à l'ADN. D'autre part, les ADN polymérases dites « alternatives » ou « spécialisées » interviennent dans la synthèse translésionnelle ou de réparation de l'ADN en cas de stress exogène (Figure 14A).

ADN polymérases répliquatives		
α , alpha	PolA	Famille B
δ , delta	PolD1	
ϵ , epsilon	PolE	

ADN polymérases alternatives		
β , beta	PolB	Famille X
λ , lambda	PolL	
μ , mu	PolM	
σ , sigma	PolS	
TdT	TDT	
ι , iota	PolI	Famille Y
κ , kappa	PolK	
η , eta	PolH	
Rev1	REV1L	
ν , nu	PolN	Famille A
θ , theta	PolQ	
ζ , zeta	PolZ	Famille B
Primpol	CCDC111	Famille AEP

B.

ADN polymérase répliquative

ADN polymérase alternative

Figure 14 : Les deux grandes familles d'ADN polymérases eucaryotes. (A) Il existe 16 ADN polymérase chez les eucaryotes, réparties en deux grandes familles : les ADN pol répliquatives et les ADN pol alternatives. (B) Les ADN polymérases ont une structure en main droite. Le site catalytique des ADN polymérases translésionnelles est plus ouvert que celui des ADN polymérases répliquatives. Tiré de (Friedberg, 2005)

1. Les ADN polymérases répliquatives

Les ADN polymérases répliquatives, Pol δ , Pol ϵ et Pol α , impliquées dans la réplication de l'ADN nucléaire appartiennent à la famille B des ADN polymérases. Pol α est une primase : elle catalyse la synthèse d'amorces d'ARN : ADN, qui seront ensuite étendues par Pol δ et Pol ϵ . Elle intervient donc au moment de l'initiation de la réplication suite à l'activation d'une origine de réplication, puis tout au long de la duplication du brin discontinu en synthétisant les amorces des fragments d'Ogazaki. Ensuite, s'il semblait acquis que Pol δ et Pol ϵ étaient respectivement responsables de l'élongation des brins discontinu et continu (Pursell et al., 2007; Shinbrot et al., 2014), une étude chez la levure vient remettre en question cette répartition des brins parentaux entre les deux ADN polymérases répliquatives. En concluant que Pol δ est responsable de l'élongation des deux brins (Johnson et al., 2015), ils relancent le débat dans le domaine (Stillman, 2015). Pol δ et Pol ϵ peuvent aussi intervenir en cas de dommages à l'ADN, soit dans de la synthèse de réparation, soit pour la synthèse dépendant du checkpoint de phase S.

Ces trois ADN polymérases répliquatives sont très fidèles et génèrent en moyenne moins d'une erreur (substitution, insertion ou délétion) toutes les 10 000 incorporations de désoxyribonucléotides (dNTP) (pour revue (McCulloch and Kunkel, 2008)), Pol α étant la moins fidèle de toutes. En effet, cette dernière ne possède pas, à l'inverse de Pol δ et Pol ϵ , d'activité de relecture, liée à leur activité exonucléase 3'-5', qui leur permet de faire « machine arrière » en cas de mésappariement et de corriger l'erreur. La fidélité de ces ADN polymérases peut aussi être influencée par la composition et l'abondance du *pool* de dNTP disponible à la fourche ainsi que de leur sélectivité pour le choix du dNTP (Kunkel, 2011).

Les ADN polymérases adoptent une structure tri-dimensionnelle (Steitz, 1998) ressemblant à une main droite (Figure 14B), où les domaines fonctionnels sont représentés par :

- La paume où l'on retrouve le domaine catalytique responsable de l'activité polymérase. Ce domaine catalyse la réaction nucleotidyl-transférase, en ajoutant les nucleotides à l'extrémité 3'-OH du brin en cours d'élongation.
- Le pouce, qui facilite l'interaction avec l'ADN et qui est important pour la processivité de l'ADN polymérase.
- Les doigts qui permettent un positionnement correct de la matrice et des nucléotides à incorporer.

2. Les ADN polymérases spécialisées

L'autre famille d'ADN polymérases est celle des ADN polymérases spécialisées. Elle comprend 13 membres, classés en différentes familles en fonction de leurs homologues de séquences et de structure. Elles ont principalement été décrites pour leurs rôles dans la réponse aux dommages de l'ADN, soit dans des mécanismes de réparation, soit dans un mécanisme de tolérance des dommages, la synthèse translésionnelle. La synthèse translésionnelle consiste en un échange des ADN polymérases à la fourche. En effet, les ADN polymérases répliquatives sont caractérisées par un site catalytique étroit et rigide. D'un côté, cela permet une grande sélectivité des nucléotides participant à leur grande fidélité, mais d'un autre côté cela les empêche de continuer l'élongation face à une base modifiée, une structure secondaire de l'ADN, ou toute autre distorsion. Ce sont alors les ADN polymérases translésionnelles qui viennent les remplacer à la fourche grâce à leur site catalytique plus ouvert et plus souple (Figure 14B), permettant une élongation face à ces barrières de fourche de réplication. Si elles peuvent respecter la synthèse d'ADN correcte dans leur activité de translésion, leur taux d'erreur face à de l'ADN non endommagé est important (Friedberg, 2005). Cependant de plus en plus d'évidences suggèrent aujourd'hui que plusieurs d'entre elles interviennent aussi dans la réplication du génome, en absence de stress exogène (Bergoglio et al., 2013; Bhat et al., 2013; Bétous et al., 2013).

Mes premiers travaux de thèse ont permis de montrer que l'une d'entre elle, l'ADN polymérase Thêta (Pol θ), jouait un rôle tout particulier dans la réplication en absence de lésions exogènes en régulant le *timing* de réplication.

B. L'ADN polymerase Thêta

1. Structure

L'ADN polymérase thêta (Pol θ) appartient à la famille A des ADN polymérases. Chez l'Homme, c'est une protéine de grande taille (290 kDa), codée par le gène *POLQ*, possédant trois grands domaines (Figure 15A). En C-terminal (C-ter) se trouve le domaine polymérase contenant des motifs caractéristiques de la famille A des ADN polymérases, ainsi que cinq motifs qui lui sont spécifiques, les inserts 1, 2 et 3 et les boucles exo1 et exo2 (Hogg et al., 2011; Seki et al., 2004; Zahn et al., 2015). L'insert 1, situé à l'extrémité du « pouce » de Pol θ , favorise la processivité de l'enzyme en interagissant directement avec le substrat d'ADN. Les inserts 2 et 3, situés de part et d'autre de la « paume » de Pol θ , augmentent l'efficacité de polymérisation à partir d'ADN non endommagé et sont nécessaires pour ses capacités

transléesionnelles. Récemment la mise en évidence de la structure du domaine polymérase par cristallographie (Zahn et al., 2015) a permis de découvrir les deux autres motifs Exo1 et Exo2 (Figure 15B). Ils sont tous les deux situés dans un domaine présentant une structure caractéristique d'un domaine exonucléase (*exonuclease-like*), situé en N-terminal (N-ter) du domaine polymérase. Mais l'absence de certains résidus catalytiques nécessaires à cette fonction suggèrent que Polθ ne présente pas d'activité exonucléase bien que l'ADN polymérase θ purifiée à partir de cellules HeLa présente une activité 3'-5' exonucléase (Maga et al., 2002). La proximité et la position en N-ter des boucles Exo1 et Exo2 ont conduit Zahn *et al.* à supposer qu'elles pouvaient jouer un rôle dans l'interaction avec les domaines central ou hélicase de Polθ.

En effet, en N-terminal de la protéine, de façon unique dans le monde des ADN polymérases, Polθ possède un domaine de la superfamille 2 des hélicases dont la fonctionnalité n'a pas encore été démontrée *in vitro*. Ces deux régions sont séparées par un important domaine central qui lui est spécifique et aux fonctions encore inconnues (Seki et al., 2003).

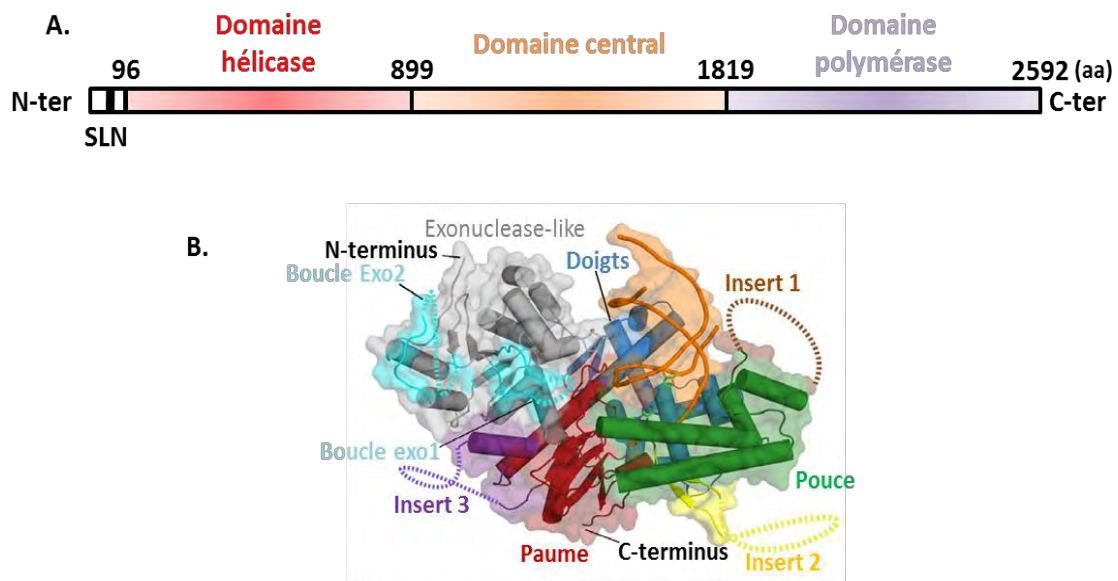


Figure 15 : Structure de l'ADN polymérase humaine. (A) Polθ est composée de trois grands domaines : hélicase, central et polymérase, de l'extrémité N-terminale à l'extrémité C-terminale. (B) Structure du domaine polymérase de Polθ mis en évidence par cristallographie. Tiré de (Zahn et al., 2015).

2. Fonctions

La conservation de Polθ au cours de l'évolution, comme en témoigne la présence d'orthologues chez plusieurs métazoaires (pour revue (Yousefzadeh and Wood, 2013)) suggère un rôle physiologique important et qui était encore mal défini au début de ma thèse.

a. *Maintien de l'intégrité du génome et réparation*

Bien qu'appartenant à la famille A des ADN polymérases dites fidèles, Polθ, de la même façon que les ADN polymérases translésionnelles de la famille Y, est infidèle *in vitro* face à une matrice d'ADN non endommagée (Arana et al., 2008; Seki et al., 2004). De nouveau, ces travaux sont en contradiction avec l'étude de Maga *et al.* qui conclut que Polθ purifiée à partir d'extraits cellulaires est une enzyme fidèle. Ces divergences peuvent s'expliquer par d'éventuels cofacteurs de Polθ qui lui confèreraient des capacités de relecture, améliorant ainsi sa fidélité. Elle est capable de catalyser une synthèse translésionnelle face à des lésions de l'ADN de type oxydatives, telles que les sites abasiques ou thymine glycol, en incorporant une base face à la lésion puis en continuant la synthèse d'ADN sur quelques nucléotides (Seki et al., 2004; Yoon et al., 2014). Par ailleurs, chez le poulet ou chez *C. Elegans* elle joue un rôle dans des mécanismes de réparation, notamment dans la réparation par excision de base en réponse à ces dommages oxydatifs, mais sa participation semble toutefois mineure (Asagoshi et al., 2012; Yoshimura et al., 2006). De plus, chez la souris invalidée pour *POLQ*, le taux de micronoyaux spontanés est augmenté dans les érythrocytes témoignant d'une augmentation des cassures double-brin spontanées en absence de Polθ (Shima et al., 2004). Parallèlement, la déplétion de cette ADN polymérase chez la souris ou des lignées de cellules humaines tumorales induit une augmentation de la sensibilité aux radiations ionisantes, ou à d'autres agents génotoxiques induisant des cassures de l'ADN (Goff et al., 2009; Higgins et al., 2010a) suggérant un rôle dans la réparation des cassures double-brin (CDB). Et en effet, un ensemble de travaux très récents ont montré que Polθ participe à une voie de réparation des CDB par recombinaison non homologue, le NHEJ alternatif (pour « *Non Homologous End Joining* ») ou MMEJ (pour « *Microhomology Mediated End Joining* ») chez les mammifères (Ceccaldi et al., 2015a; Kent et al., 2015; Mateos-Gomez et al., 2015; Yousefzadeh et al., 2014; Zahn et al., 2015), venant corroborer des résultats obtenus quelques années plus tôt chez la drosophile (Chan et al., 2010) ou *C.elegans* (Koole et al., 2014; Roerink et al., 2014).

b. *Rôle dans le NHEJ alternatif*

Les CDB représentent une des lésions les plus toxiques pour l'organisme, car pouvant engendrer la mort de la cellule par voie apoptotique ou une très forte instabilité génétique avec des délétions, insertions ou translocations. Ces événements génétiques sont en particulier fréquemment retrouvés dans les cancers. Il existe plusieurs voies de réparation de ces cassures : la recombinaison homologue (RH), le NHEJ classique (c-NHEJ) ou alternatif (alt-

NHEJ), et la réparation par SSA (pour « *Single Strand Annealing* ») (Figure 16). Les deux majoritaires sont la HR et le NHEJ classique. Brièvement la RH est un processus de réparation généralement fidèle des CDB. Il repose sur l'invasion de brin de la chromatide sœur, qui sert alors de modèle pour la synthèse du brin cassé. Le c-NHEJ est aussi en lui-même un processus de réparation fidèle, en particulier dans le cas de cassures à extrémités franches, mais peut engendrer des délétions de quelques nucléotides si les extrémités de la CDB sont cohésives, présentant des extrémités simple-brin très courtes (quelques nucléotides) (Figure 16). Dans ce processus, les deux extrémités de la cassure reconnues par Ku70/80 sont recollées entre elles par une ligase, la ligase IV. Enfin, le SSA et le alt-NHEJ reposent sur la recherche d'homologies entre les deux extrémités de la cassure : le SSA sur de longues homologies et le alt-NHEJ sur de courtes ou micro-homologies de séquences. Ces deux derniers mécanismes de réparations des CDB sont certainement les plus mutagènes, engendrant respectivement de longues délétions, ou des délétions et insertions de séquences (Pour revue (Ceccaldi et al., 2015a; Kent et al., 2015; Mateos-Gomez et al., 2015; Yousefzadeh et al., 2014; Zahn et al., 2015)).

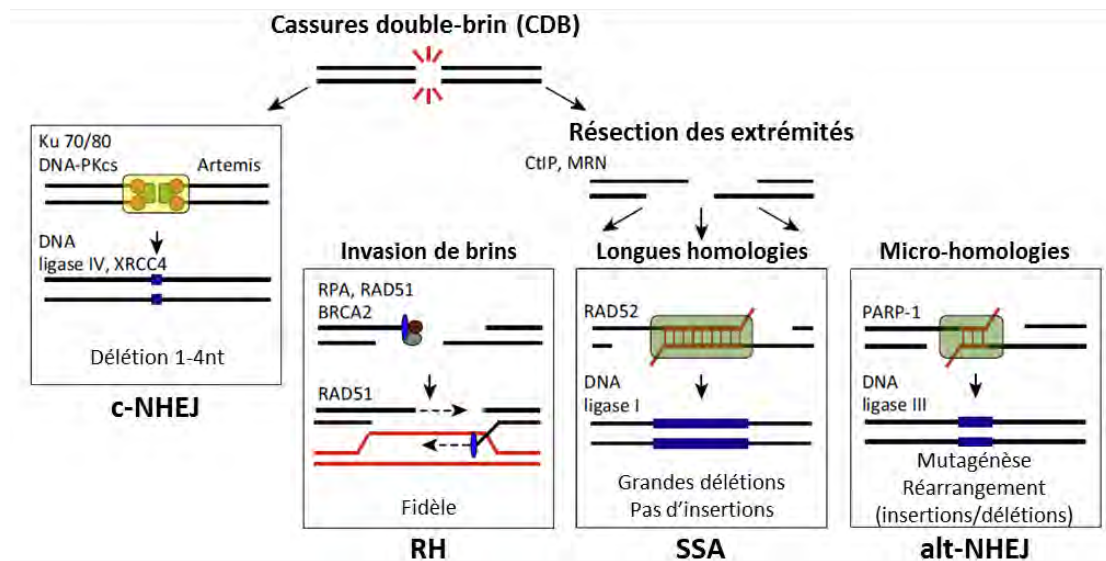


Figure 16 : Mécanismes de réparation des cassures double-brin.

Adapté de (Ceccaldi et al., 2015b)

L'implication de Polθ dans le alt-NHEJ chez les mammifères a d'abord été mise en évidence chez la souris par l'équipe de R. Wood (Yousefzadeh et al., 2014). Partant du constat que les cellules *polq*^{-/-} présentaient une hypersensibilité aux agents génotoxiques provoquant des CDB, les auteurs ont pu mettre en évidence son implication dans le alt-NHEJ. En effet, l'absence de Polθ affectait la réparation de CDB mimées par des substrats d'ADN

double-brin si les extrémités cohésives présentaient une extrémité 3' sortante longue (45nt). Si cette extrémité était courte, (6nt), ce n'était pas le cas. A l'inverse, les MEF déplétées pour Ku70, acteur du c-NHEJ, présentent un défaut dans la réparation des CDB lorsque l'extrémité 3' est courte. D'un point de vue mécanistique, Polθ, *in vitro*, est capable de catalyser l'élongation en 3' à partir d'un ADN simple-brin dès lors qu'une hybridation intra ou inter brin est possible, même sur un nombre réduit de nucléotides. Ainsi, elle participerait au alt-NHEJ en permettant l'hybridation des deux extrémités de la CDB à partir de l'appariement d'un petit nombre de bases. L'étude de Kent *et al.* vient confirmer et compléter ces résultats (Kent et al., 2015). En effet, *in vitro*, Polθ synthétise de l'ADN uniquement sur des substrats double-brin présentant une extrémité 3' flottante d'au moins 20 nucléotides (nt) capable de s'hybrider *via* des micro-homologies de séquences d'au moins 4 nt. Si l'hybridation ne se fait que sur 2 nt, l'élongation reste possible mais son efficacité est très limitée. Ce substrat mime une extrémité de CDB partiellement résectée, pour se rapprocher d'un substrat existant *in vivo* après résection par CtIP ou Mre11. Polθ est donc un acteur du MMEJ. La position de ces séquences complémentaires dans les extrémités 3' importe peu : qu'elles se situent à l'extrémité ou au sein de la queue flottante ne change pas les capacités d'élongation de Polθ (Figure 17). En revanche, la qualité de ces micro-homologies influence l'efficacité du MMEJ, qui est facilitée si elles sont riches en GC. Enfin, Polθ participe aussi au MMEJ indépendamment de son activité polymérase en autorisant la formation de la synapse de réparation. Par dimérisation, elle rapproche spatialement les deux extrémités de la CDB, qu'il y ait micro-homologie ou non. Ensuite, si c'est le cas, elle dirige l'appariement par micro-homologies de séquences. Enfin, son activité polymérase entrera en jeu pour synthétiser l'ADN manquant (Figure 17). Bien que les études de cristallographie n'aient pas pu le mettre en évidence, l'insert 2 spécifique de Polθ (Figure 15B) semble nécessaire à la dimérisation de la protéine et permet l'affinité de la polymérase pour l'ADN partiellement résecté (Kent et al., 2015; Zahn et al., 2015).

En plus d'être un membre actif de la voie du alt-NHEJ, Polθ inhibe en parallèle l'autre voie de réparation des CDB après résection, la RH, en limitant le recrutement de Rad51 au niveau des cassures. En effet, sa déplétion induit une accumulation des foyers Rad51 et une augmentation de la réparation des CDB par RH après radiations ionisantes (RI) (Ceccaldi et al., 2015a; Mateos-Gomez et al., 2015). Cette régulation de la HR par Polθ implique le domaine ATPase de la protéine ainsi qu'un de ses sites de liaison à Rad51. En effet, après extinction du gène *POLQ*, les expériences de complémentation par des mutants de ces domaines ne permettent pas de restaurer le niveau basal de foyers Rad51 ou de RH observés

dans des conditions contrôle, toujours après RI. En revanche, c'est le cas quand les cellules sont complétées par la forme sauvage de la protéine montrant bien le rôle spécifique de Polθ, ou par un mutant déplété pour le domaine polymérase. Il semble donc que l'activité polymérase ne soit pas impliquée dans la compétition avec la RH. En interagissant directement avec Rad51, Polθ empêche la formation de la D-loop par Rad51 au moment de l'invasion de la chromatide sœur, inhibant ainsi la RH (Ceccaldi et al., 2015a). Polθ joue donc un rôle dans une voie de réparation des CDB complémentaire et en compétition avec la RH. Ainsi, dans un contexte RH déficient, la présence de Polθ est essentielle. D'une part, la déplétion de Fancd2 et de Polθ chez la souris est létal pour l'embryon. D'autre part, en absence de Polθ des cellules déficientes pour la RH ne survivent pas aux traitements par des agents qui inhibent la réplication. Ces résultats suggèrent donc que dans le cadre d'une thérapie du cancer, une stratégie de létalité synthétique peut être envisagée sur des tumeurs déficientes pour la RH, en couplant une chimiothérapie ciblant l'ADN à une inhibition de Polθ.

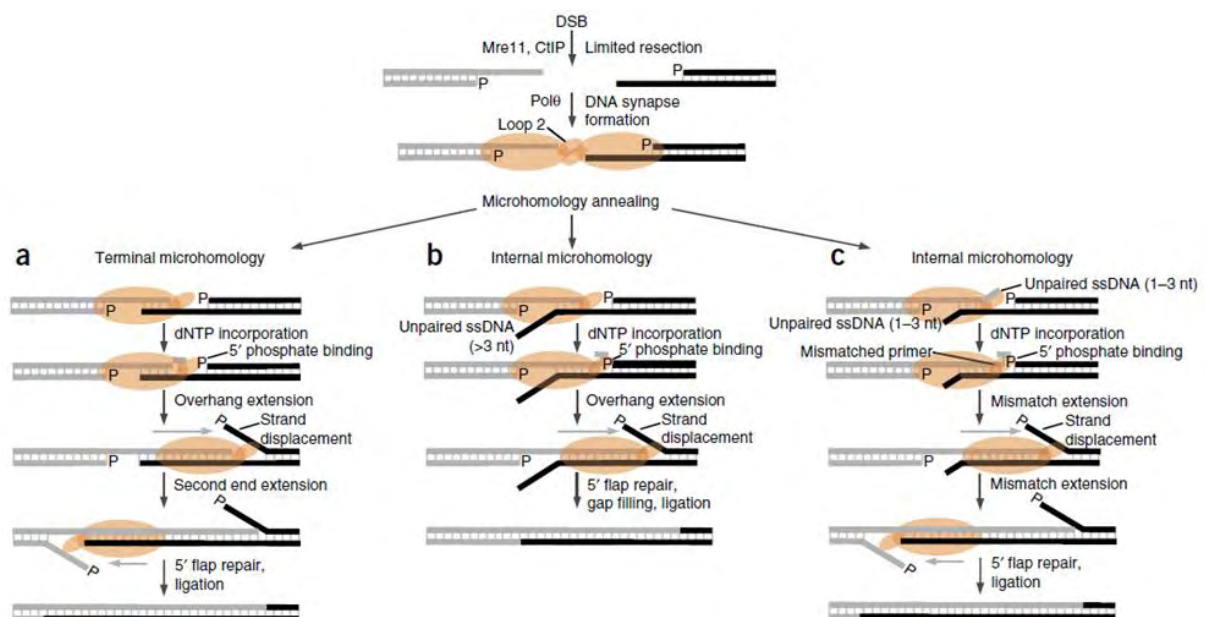


Figure 17 : Modèle pour le rôle de Polθ dans le MMEJ. (tiré de (Kent et al., 2015)). Suite à la résection des extrémités 5' de la CDB par CtIP et Mre11, le dimère de Polθ provoque la formation de la synapse de réparation en recrutant les deux extrémités de la cassure. Elle permet ainsi l'hybridation des deux queues 3' flottantes au niveau de microhomologies, qui peuvent être situées à l'extrémité de la queue (a), à l'intérieur de la séquence (c), ou bien un mélange des deux (b). Elle synthétise ensuite l'ADN manquant, en déplaçant l'extrémité 5'. La réparation est enfin résolue par la réparation des queues flottantes 5' ainsi générées, du gap-filling si nécessaire (b) et la ligation des fragments d'ADN néosynthétisés avec la matrice.

c. *Autres fonctions décrites*

Une des fonctions de Polθ largement explorée est son implication potentielle dans l'hypermutation somatique (SHM) des immunoglobulines (IgG) ou la commutation de classe des IgG. La question du rôle de Polθ dans la maturation des IgG s'est posée suite au constat d'une expression forte de l'ARN messager (ARNm) de *POLQ* dans les tissus lymphoïdes, et d'une augmentation de cette expression suite à l'induction, en particulier à des stades de la différenciation des lymphocytes B (L_B) où la SHM est élevée (Kawamura et al., 2004; Zan et al., 2005). L'ensemble des études à ce sujet sont finalement assez contradictoires quant à l'importance que pourrait jouer cette ADN polymérase dans ce domaine. En effet, d'un côté, la perte de *POLQ* dans les plaques de Peyer entraîne une diminution de 60 à 80% de la SHM, parallèlement à un appauvrissement du spectre de mutation (Zan et al., 2005). D'un autre côté, dans les cellules B issues de la rate 14 jours après immunisation de souris déplétées pour *POLQ*, la diminution de la SHM est très légère, mais significative (Masuda et al., 2006). Ces différences observées peuvent s'expliquer par l'origine des cellules B. En effet, les plaques de Peyer sont exposées de façon chronique aux antigènes environnementaux, générant un taux basal de mutations élevé. Les cellules B prélevées dans la rate témoignent, elles, d'une réponse immunitaire à court terme. De plus les méthodes d'amplification du locus témoin soumis à la SHM, ou le fond génétique des souris différent, peuvent aussi expliquer ces différences. Une coopération de Polθ avec Polη dans la SHM a aussi été envisagée (Masuda et al., 2007). Mais de nouveau, une étude contradictoire vient minimiser ce rôle éventuel (Martomo et al., 2008). Par ailleurs, dans les cellules DT40 de poulet, Polθ et Polv coopèrent dans un autre mécanisme de diversification de la partie variable des anticorps, la conversion des gènes IgG, qui repose sur une recombinaison homologue entre deux parties du chromosome. Enfin, Polθ intervient dans une sous-classe de la permutation de classe des immunoglobulines, en lien avec son rôle dans le alt-NHEJ (Yousefzadeh et al., 2014). Au final, si Polθ intervient dans la diversification des IgG, il semble bien que son rôle ne soit pas majeur, mais peut-être spécifique à certaines conditions qui resteraient à définir.

Enfin, une étude décrit un rôle d'un analogue de Polθ, le gène *TEBICHI*, dans le contrôle du cycle cellulaire et la différenciation des cellules du méristème chez la plante (Inagaki et al., 2006). En effet, chez un mutant hélicase de la protéine, de nombreux défauts morphologiques sont observés sur différents organes de la plante (racine, feuilles, méristème). De plus, la différenciation du méristème est altérée, et des dérégulations du cycle cellulaire sont observées dans les premiers stades développementaux, rendant les embryons anormaux.

Enfin, chez ce mutant, la réponse aux dommages à l'ADN est constitutivement activée, en particulier les gènes impliqués dans la RH.

3. Signature dans les cancers

Dès 2004, les travaux de Kawamura et al. ont montré l'existence d'une surexpression de *POLQ* dans les tissus tumoraux par rapport aux tissus sains correspondants. Malgré des échantillons de patients très petits, cette étude faisait le lien entre surexpression et mauvais pronostic de la maladie dans le cancer du côlon (Kawamura et al., 2004). Depuis, des études à grandes échelles réalisées sur de plus grandes cohortes ont aussi mis en évidence la surexpression de *POLQ* dans de nombreuses tumeurs solides dont celles de l'estomac, du colon, mais aussi dans le cancer du sein, du poumon, les cancers ovariens ou les cancers Oto-Rhino-Laryngologiques. De plus, cette surexpression est associée à un facteur de mauvais pronostic dans le cancer du sein (Higgins et al., 2010b; Lemée et al., 2010) et du poumon (Allera-Moreau et al., 2012; Kawamura et al., 2004). C'est également le cas dans le cancer du côlon lorsque cette surexpression est associée à celle d'au moins trois gènes de l'initiation de la réplication, dont *CDC45*, *CDC6*, *CDT1*, *MCM2* et *MCM7* (Pillaire et al., 2010).

Le rôle de Pol θ dans le processus tumoral n'est pas encore identifié. Sa surexpression participe-t-elle à la genèse de la maladie ? Ou bien sa surexpression est-elle induite et sélectionnée au cours du processus tumoral ? On sait que la surexpression de cette polymérase dans des fibroblastes humains induit une perturbation du programme de réplication (vitesse des fourches de réplication ralentie), associée à une augmentation de l'instabilité génétique (foyers γ H2AX, et anomalies chromosomiques) (Lemée et al., 2010). Au cours du développement de la maladie, la surexpression de Pol θ pourrait donc, en perturbant le programme de réplication, induire une forte instabilité génétique motrice dans le développement tumoral. Ses fonctions dans la réparation des CDB pourraient permettre aux cellules surexprimant Pol θ de résister au stress réplcatif intrinsèque des cellules tumorales surtout dans le cas où les gènes de la RH sont déficients. De la même façon, il est possible qu'elle participe à la résistance aux chimiothérapies ciblant l'ADN. Les cellules surexprimant Pol θ seraient alors sélectionnées pendant les périodes de traitement, favorisant résistance ou rechutes.

Au début de ma thèse, le rôle de l'ADN polymérase Thêta était encore très peu défini. Et sa signature particulière et unique dans les cancers ont donc amené mon équipe à poser la question de son rôle physiologique, en absence de stress exogène. Dans ce cadre, les travaux

réalisés avec et sous la direction d'Anne Vidal-Fernandez ont permis de révéler un tout nouveau rôle pour Pol θ dans le *timing* de réplication, encore jamais décrit dans le monde des ADN polymérase. Nous avons tout d'abord mis en évidence que Pol θ est recrutée à la chromatine très tôt en phase G1 ; au niveau de certaines origines de réplication, comme le suggèrent les co-immunoprécipitations de Pol θ avec deux composants du pré-RC, Orc2 et Orc4. En outre, sa déplétion induit une augmentation du recrutement des hélicases réplcatives Mcm. Enfin, nous avons observé que la dérégulation de l'expression de Pol θ induit des modifications du *timing* de réplication sur l'ensemble du génome. Ces travaux ont fait l'objet d'une publication dans Nature Communications, en juillet 2014.

II. Un nouveau rôle pour l'ADN polymérase Thêta dans le *timing* de réplication

ARTICLE

Received 8 Jan 2014 | Accepted 3 Jun 2014 | Published 3 Jul 2014

DOI: 10.1038/ncomms5285

A role for DNA polymerase θ in the timing of DNA replication

Anne Fernandez-Vidal^{1,2,*}, Laure Guitton-Sert^{1,2,*}, Jean-Charles Cadoret^{3,*}, Marjorie Drac⁴, Etienne Schwob⁴, Giuseppe Baldacci³, Christophe Cazaux^{1,2} & Jean-Sébastien Hoffmann^{1,2}

Although DNA polymerase θ (Pol θ) is known to carry out translesion synthesis and has been implicated in DNA repair, its physiological function under normal growth conditions remains unclear. Here we present evidence that Pol θ plays a role in determining the timing of replication in human cells. We find that Pol θ binds to chromatin during early G1, interacts with the Orc2 and Orc4 components of the Origin recognition complex and that the association of Mcm proteins with chromatin is enhanced in G1 when Pol θ is downregulated. Pol θ -depleted cells exhibit a normal density of activated origins in S phase, but early-to-late and late-to-early shifts are observed at a number of replication domains. Pol θ overexpression, on the other hand, causes delayed replication. Our results therefore suggest that Pol θ functions during the earliest steps of DNA replication and influences the timing of replication initiation.

¹Equipe Labellisée Ligue contre le Cancer 2013 INSERM Unit 1037; CNRS ERL 5294; CRCT (Cancer Research Center of Toulouse), BP3028, CHU Purpan, Toulouse 31024, France. ²Université Paul Sabatier, University of Toulouse III, Toulouse F-31062, France. ³Institut Jacques Monod, UMR7592, CNRS and University Paris-Diderot, 15 Rue Hélène Brion, Paris, Cedex 13 75205, France. ⁴Institut of Molecular Genetics, CNRS UMR5535 and University of Montpellier, Montpellier 34293, France. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.C. (email: cazaux@ipbs.fr) or to J.-S.H. (email: jseb@ipbs.fr).

Human DNA polymerase θ (Pol θ) is a 290-kDa protein that comprises a carboxy-terminal A-family polymerase domain, which performs template-dependent and -independent DNA synthesis *in vitro*^{1,2}, a large central domain of unknown function, and an amino-terminal domain that shares sequence similarity to known helicases, although no helicase activity has thus far been demonstrated². Pol θ also exhibits deoxyribophosphodiesterase activity³ and ATPase activity in the presence of single-stranded DNA². The recombinant enzyme lacks the 3' \rightarrow 5' exonuclease activity that is a common feature of low fidelity DNA polymerases and generates single base substitutions at an average rate comparable to the inaccurate Y family of DNA polymerases^{2,4,5}. However, an active fraction of Pol θ purified from HeLa cells was shown to be capable of high fidelity of DNA polymerization⁶. This suggests that the enzyme may be accurate in a physiological context, perhaps through cooperation with proofreading co-factors. In spite of the intensive work aimed at understanding the cellular function(s) of this polymerase carried out over the past decade, a clear picture has yet to emerge as mostly species and cell type-specific functions have been reported. For example, studies performed in mice have suggested a role for Pol θ in somatic hypermutation of immunoglobulin genes^{7–9}, but growing evidence indicate that such a function may be minor¹⁰. Pol θ has also been proposed to function in DNA repair. Indeed, this enzyme acts as a backup DNA polymerase during base excision repair in chicken DT40 cells¹¹ and nematodes¹², but this functional observation has not been clearly demonstrated in mammals^{13–15}. Studies in *Drosophila* have also suggested a role for Pol θ in the repair of double-stranded breaks through the synthesis of DNA during alternative microhomology-mediated end joining^{16,17}. This potential DNA repair function could explain why mouse Pol θ -deficient bone marrow stromal cells and human Pol θ -depleted tumours exhibit increased sensitivity to ionizing radiation^{13,14}. This role in the defense against ionizing radiation-induced DNA damage might also be due to the ability of the polymerase to perform translesion synthesis (TLS) past abasic sites and thymine glycol lesions, which are major products of reactive oxygen species-induced DNA damage and strong blockers to most other A-family as well as B-family replicative DNA polymerases^{18,19}. The polymerase domain of Pol θ seems to be sufficient for this process, which relies on three insertion loops not present in other A-family DNA polymerases¹⁸. Moreover, Pol θ has the rare ability to catalyse both the insertion and the extension steps necessary to bypass the DNA lesion^{19–21}. Nevertheless, a clear understanding of Pol θ physiological function and mechanism of action, especially under normal growth conditions, is still lacking.

By conducting large-scale studies of the expression of DNA replication factors in patients with colon, breast and lung cancer, we recently uncovered major alterations in the expression pattern of *POLQ*, the gene encoding Pol θ , and discovered that excess *POLQ* correlates with genetic instability and poor prognosis^{22–24}. This is, to our knowledge, the strongest link between a negative clinical outcome and the expression of a DNA polymerase²⁵. The unique *POLQ* expression signature among replicative DNA polymerases and other specialized TLS and repair DNA polymerases—combined with the fact that Pol θ is the only human DNA polymerase that contains a helicase domain—prompted us to speculate that Pol θ might have an essential function in unstressed cells. To determine the role of Pol θ during normal DNA replication in human cells, we asked whether either loss or excess of Pol θ could affect the DNA replication programme under physiologic growth conditions. These experiments uncovered an unexpected requirement for Pol θ for proper minichromosome maintenance (Mcm) loading during G1 and the timing of DNA replication at a subset of domains during S phase.

Results

Association of Pol θ with chromatin in early G1. We first studied in detail Pol θ recruitment onto chromatin during the cell cycle by using unstressed human RKO cells under physiological growth conditions (5% oxygen concentration). Cells were synchronized in pro-metaphase with nocodazole, collected by mitotic shake-off, released into the cell cycle and then analysed at various times throughout M, G1 and S phases. We checked cell cycle progression and S phase entry by bromodeoxyuridine (BrdU) incorporation, propidium iodide staining and fluorescence-activated cell sorting analysis (Fig. 1, left panel). Cell extracts were then fractionated and chromatin fractions were subjected to immunoblotting (Fig. 1, right panel). Unexpectedly, detectable Pol θ was found in the chromatin fraction as cells exited mitosis. Its level in this fraction then increased greatly in early G1 and persisted until S phase. As G1 is the phase where the replication origins are set up and the pre-replication complexes (pre-RCs) formed^{26,27}, we compared the timing of Pol θ recruitment with the kinetics of pre-RC assembly. As expected, Origin recognition complex subunit 2 (Orc2) and cell division control protein 6 (Cdc6) were associated with chromatin in mitosis and remained there as cells entered G1. Next, Cdt1 (Cdc10 protein-dependent transcript 1) was recruited, followed by the recruitment of Mcm helicases and the downregulation of Cdc6 (ref. 28). As also expected, when cells entered S phase the enzymes involved in DNA synthesis, including the replicative A-family DNA polymerases Pol α and Pol δ , were recruited and Cdt1 released. Strikingly, Pol θ was found to be recruited to chromatin just before Mcm. All of these observations were confirmed using the Cdk1 inhibitor RO-3306 as an alternative mode of cell synchronization (Supplementary Fig. 1). Under these conditions, Pol θ was recruited to chromatin concomitantly with Mcm. These results appear to be the first evidence of a DNA polymerase being loaded onto chromatin in early G1 in the absence of external stress.

Pol θ interacts with Orc2 and Orc4. Assembly of the pre-RC complex starts with the binding of Orc. This then serves as a platform for the loading of Cdc6, Cdt1 and Mcm2–7. Our observation that Pol θ associates with chromatin around the time of pre-RC formation in early G1 prompted us to analyse whether Pol θ may interact directly with the Orc platform. First, a Pol θ antibody was used to immunoprecipitate Pol θ from cell lysates. Orc2 and Orc4 antibodies were then used to test for the presence of Orc2 and Orc4 by western blotting (Fig. 2a). We identified Orc2 and Orc4 in Pol θ immunoprecipitates and these interactions were confirmed by performing immunoprecipitation with different Orc2 and Orc4 antibodies followed by western blot analysis of the immunoprecipitated Pol θ (Fig. 2a). To the best of our knowledge, Orc2 and Orc4 are the first identified proteins that can interact with Pol θ . These data suggest that Pol θ could bind to replication origins.

Pol θ regulates Mcm loading onto chromatin in G1. Our observations that Pol θ interacts with Orc2 and Orc4 prompted us to explore whether Pol θ is required for pre-RC formation. Therefore, we analysed pre-RC assembly in G1 in the absence of Pol θ (Fig. 2b). Small interfering RNA (siRNA)-mediated *POLQ* gene silencing was performed in RKO cells with a near-complete depletion of Pol θ (Fig. 2b, right panel). Pol θ depletion did not affect cell cycle progression even after long-term downregulation (Supplementary Fig. 2A) and did not trigger the activation of the DNA damage signalling (Supplementary Fig. 2B). Mock and Pol θ -depleted cells were synchronized in G2/M with nocodazole and then released into fresh medium and collected in G1 270 min

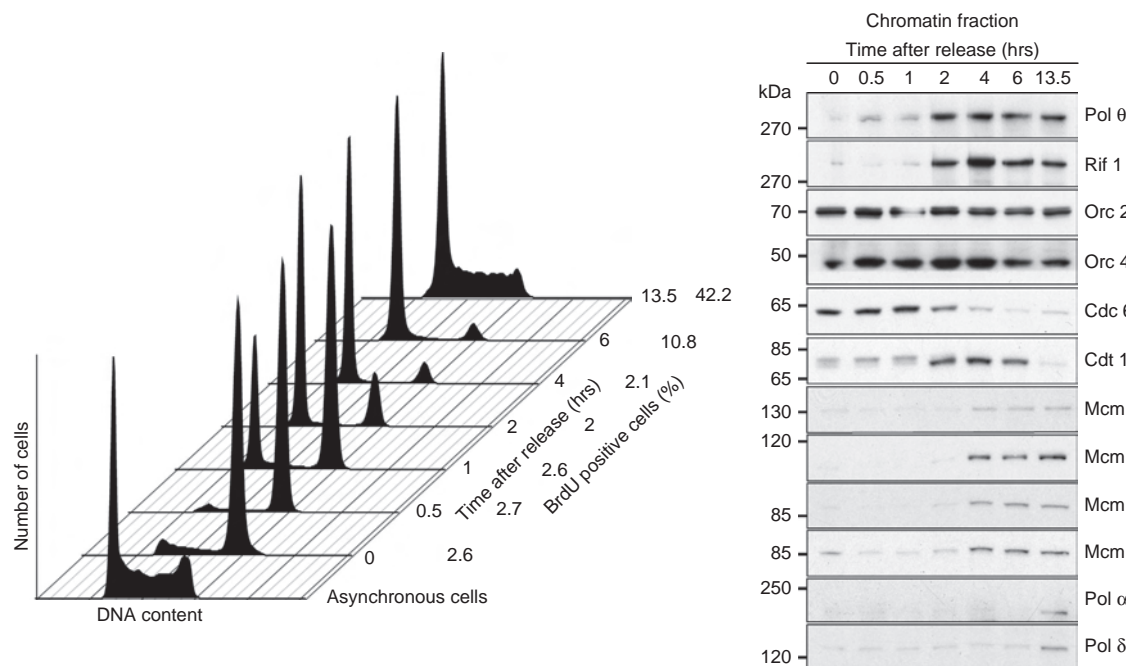


Figure 1 | Pol θ is recruited to chromatin in early G1 in human cells. Cells of the human RKO line were synchronized in mitosis by nocodazole treatment (0.1 μ M, 20 h) and then released into fresh medium for the indicated times (0, 30 min, 1, 2, 4, 6 h and 13 h 30 min). Cell cycle progression was followed by fluorescence-activated cell sorting analysis after BrdU incorporation and propidium iodide staining (left panel). The recruitment of Pol θ onto chromatin through the cell cycle was analysed by western blotting after subcellular fractionation (right panel). The time course of chromatin binding for Rif1, the pre-replication complex and the replicative DNA polymerases α and δ were also checked (right panel). Equivalent cell number fractions were loaded. This figure is representative of three independent experiments.

after release (Fig. 2b, left panel). Cell extracts were fractionated and immunoblotting of the resulting fractions revealed that when Pol θ was downregulated the binding of Orc2, Cdc6 and Cdt1 to chromatin remained unchanged (Fig. 2b, right panel). In contrast, the association of Mcm2, Mcm4 and Mcm7 with chromatin was significantly increased in the absence of Pol θ . These observations were confirmed in a separate *POLQ* siRNA experiment where cell synchronization was achieved using the Cdk1 inhibitor RO-3306 (Supplementary Fig. 3, right panel). Mcm accumulation was not due to a cycle effect, as Pol θ depletion did not lead to a significant modification of cell cycle progression (Fig. 2b, left panel, and Supplementary Fig. 3, left panel). Therefore, collectively these results show that Pol θ plays an important role in controlling the association of Mcm with chromatin in G1. As the Mcm complexes are known to be initially recruited and enriched at replication origins during G1 and as Pol θ interacts with Orc2 and Orc4, this suggests that Pol θ contributes to the proper formation of pre-RC complexes directly at replication origins.

Pol θ does not control origin density. We next hypothesized that the increased Mcm loading seen in G1 in Pol θ -deficient cells may lead to an increased number of origins, including active origins. To test this, we quantified the density of active origins using single molecules analysis by DNA combing, after sequential pulse-labelling of control, and Pol θ -depleted cells with the nucleotide analogues IdU and CldU²⁹. We first measured the inter-origin distances (IODs) at the genome-wide level, which revealed the number of active origins within replication domains. We did not observe a significant change in IODs in Pol θ -depleted cells relative to control cells (Fig. 3a). We then monitored global fork density, which reflects the number of active origins both within and between different replication domains. Again, we found that the global fork density, measured as the

number of forks divided by the total length of DNA fibres and normalized by the percentage of cells in S phase, remained unchanged following Pol θ downregulation (~ 0.9 fork per Mb) (Fig. 3b). As a positive control, we verified that upon replication stress (low doses of Hydroxyurea), RKO cells showed inhibition of distant origins, as revealed by a decreased fork density (Supplementary Fig. 4). From these experiments we concluded that the increased Mcm loading in G1 in Pol θ -depleted cells did not result in increased origin firing during S phase, suggesting that the loss of Pol θ did not lead to the formation of additional pre-RCs. This conclusion is supported by our observations that loading of the pre-RC proteins Cdc6 and Cdt1 was not significantly increased following Pol θ depletion (Fig. 2b, right panel).

Pol θ depletion affects the replication timing. As the increased level of Mcm accumulation onto chromatin in the absence of Pol θ did not result in a change in origin density, we investigated whether the timing of origin activity could be affected in Pol θ -depleted cells. Activation of replication origins is controlled by a specific timing programme established in early G1 phase at the so called ‘timing decision point’ (TDP)^{30,31}. Replication timing is a highly stable cell type-specific process whose mechanistic basis remains poorly understood in vertebrates^{32–34}. The data presented in Fig. 1 show that the timing of the association of Pol θ with chromatin is close to that of the TDP. Moreover, it coincides with the loading of Rap-interacting-factor-1 (Rif1) (Fig. 1, right panel), recently discovered as the first molecular regulator of replication timing^{32,34,35}. We therefore speculated that Pol θ could contribute to the determination of the replication timing profile. This issue was addressed by analysing the timing of replication domains in control and Pol θ -depleted cells (Fig. 4a) at a genome-wide level. Cells were pulse-labelled with

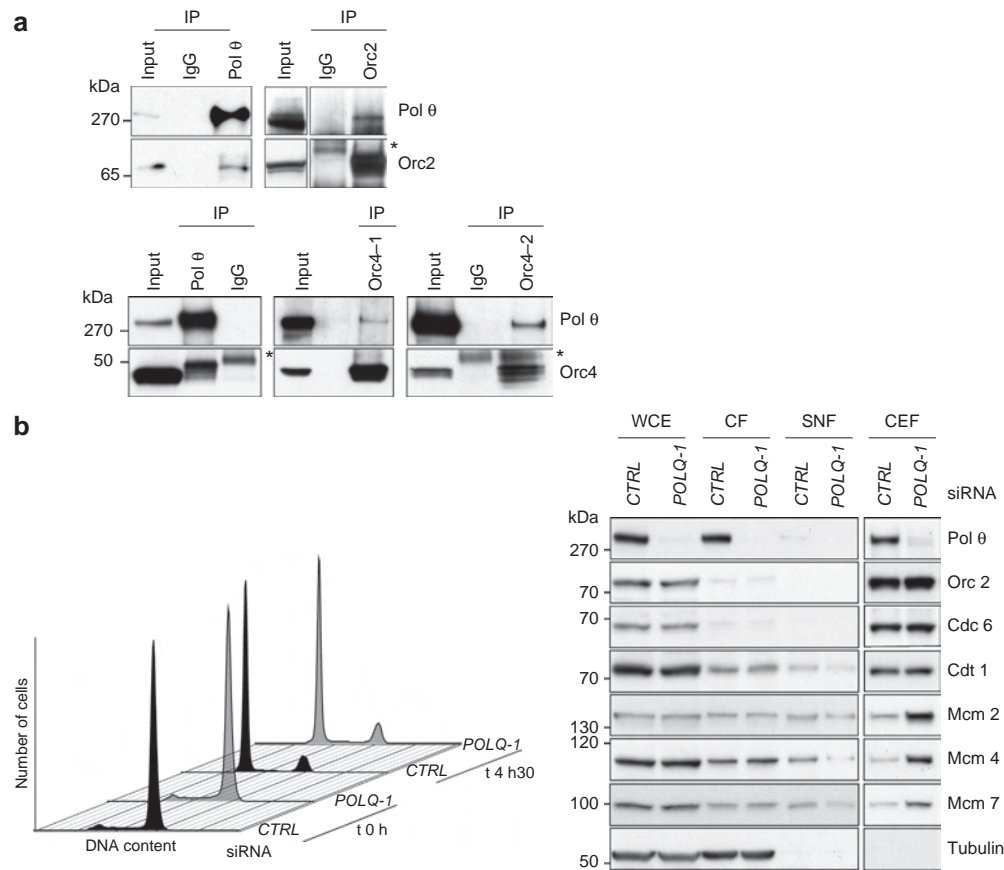


Figure 2 | Pol θ interacts with Orc2 and Orc4, and controls Mcm loading onto chromatin in G1. (a) Pol θ interacts with Orc2 and Orc4. RKO cell extracts were immunoprecipitated with Pol θ , Orc2, or Orc4 antibodies (Orc2-1, Orc4-1 and Orc4-2) and analysed by western blotting using anti-Pol θ , Orc2 or Orc4-2 antibodies; asterisk (*) corresponds to IgG. (b) Thirty hours after Control (CTRL) or POLQ-1 siRNA transfection, RKO cells were synchronized in G2/M with nocodazole (0.1 μ M, 20 h), released into fresh medium, and then collected 270 min later in G1. Cell cycle analysis of mock- (CTRL siRNA) and Pol θ - (POLQ-1 siRNA)-depleted cells was performed by propidium iodide staining and fluorescence-activated cell sorting analysis (left panel). Cell fractionation was performed and processed for immunoblotting with the indicated antibodies (right panel). All blots are representative of three independent experiments. WCE, whole cell extract; CF, cytoplasmic fraction; SNF, soluble nuclear fraction; CEF, chromatin-enriched fraction. Orc2 and Tubulin serves as fractionation and loading controls. Uncropped blots shown in (a) are shown in Supplementary Fig. 8.

BrdU and separated into early and late S-phase fractions by flow cytometry as previously described^{31,36} (see Methods). To examine replication timing, a CGH algorithm developed by the AGILENT company was used. It generates segments corresponding to both the early replication domain when the log-ratio is positive and conversely the late replication domain when the log-ratio is negative (Fig. 4b and Supplementary Fig. 5). The loss of Pol θ resulted in a modification in replication timing of \sim 5% of the human genome with an equivalent number of early-to-late (EtoL) and late-to-early (LtoE) switches (a global analysis is shown in Supplementary Fig. 5; one example from two independent experiments is presented in Fig. 4b, with black and grey arrows indicating EtoL and LtoE changes, respectively). These results clearly show that the replication timing of a subset of domains is modified following Pol θ depletion. The observation that Pol θ and Rif1 showed similar profiles of association onto chromatin after mitosis and in G1 (Fig. 1 and Supplementary Fig. 1A) further supports the idea that Pol θ contributes towards setting up the TDP. Importantly, Pol θ depletion affected neither the global level of Rif1 nor its recruitment onto chromatin in G1 (Fig. 4a and Supplementary Fig. 6A,B). Furthermore, in contrast to Pol θ -depleted cells, Rif1-depleted cells display no change in pre-RC formation, longer IOD and changes in replication timing that occur in middle S-phase domains³⁴. Collectively, these

observations suggest that similar to Rif1, Pol θ plays a role in establishing the TDP but that the mechanisms of action of Pol θ and Rif1, as well as their targeted sub-domains, are distinct. For both control and Pol θ -depleted cells, the range of the sizes of the replication domains varied from \sim 10 kb to 1.5 Mb, with several domains larger than 2 Mb (up to 8 Mb), in close agreement with previous reports in human and mouse cell lines³¹ (Fig. 4c). Moreover, late domains were always significantly larger than early domains for the control cells (Fig. 4c, P -value = 1.418×10^{-10}), as previously observed in mammalian cell lines³¹. This could be due to the acceleration of replication forks or alternatively to increased activation of stochastic origins in late S-phase, as has already been suggested³⁷. Strikingly, in contrast to control cells the sizes of the late and early domains were not significantly different in the Pol θ -depleted cells (Fig. 4c) and were both significantly reduced compared with control cells (early CTRL siRNA/early POLQ siRNA: P -value = 1.34×10^{-9} ; late CTRL siRNA/late POLQ siRNA: P -value < 2.2×10^{-16}). We observed 240 EtoL regions (early in control cells to late in Pol θ -depleted cells) corresponding to a delay of the replication timing in Pol θ -depleted cells in comparison with the control cells, and 244 LtoE regions (late in control cells to early in Pol θ -depleted cells). These regions where the timing is modified display an equivalent average size (Fig. 4d) and show similar

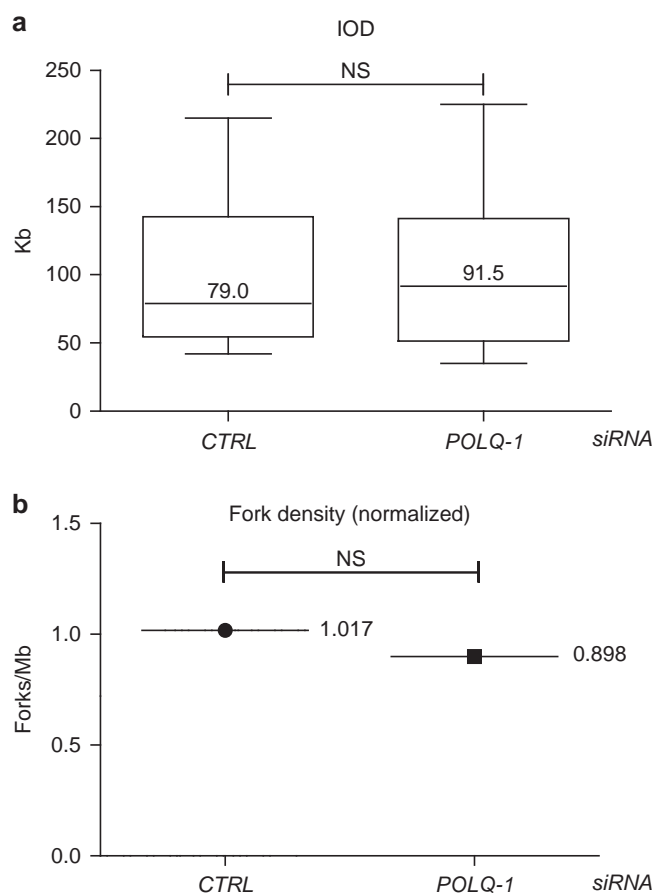


Figure 3 | Pol θ depletion does not affect the fork density and the inter-origin distances. DNA replication was monitored on single molecules spread by DNA combing in mock- (CTRL siRNA) and Pol θ - (POLQ-1 siRNA) depleted cells after sequential pulse-labelling with IdU and CldU. Inter-origin distances (**a**) and fork density (**b**) were analysed. The non-parametric test of Mann-Whitney was applied to compare data sets from control and Pol θ -depleted cells.

molecular characteristics that have been observed during cell differentiation by Hiratani *et al.*³¹

Characterization of the modified domains. The domains in which replication timing was changed in the absence of Pol θ were then characterized. Mammalian chromosomes are organized in alternative AT- and GC-rich sequences that are rich and poor in Line-1 transposable elements, respectively³⁸. Several studies have previously reported a strong positive correlation between Line 1 content, GC content, gene density and early replication^{33,39}. Such a correlation was also detectable in RKO control cells (Fig. 5a–c). Interestingly, the loss of Pol θ induced EtoL and LtoE domains that presented an intermediate Line-1 content (Fig. 5a), gene density (Fig. 5b) and GC content (Fig. 5c), similar to what was observed during cell differentiation by Gilbert and colleagues³¹. Recently, it was demonstrated that G-quadruplex (G4), a G-rich DNA sequence that can form a four stranded structure, is associated with DNA replication origins^{40,41}. Despite the absence of consensus sequences that could define replication origins, the presence of these G4 motifs may play a critical role in their localization and/or activation. As the probability of containing G4 motifs is increased in GC-rich sequences, the measures of G4 density were normalized with the GC content. As has been previously shown in different human

and mouse cell lines^{40,41}, we confirmed here that early replicating domains are enriched in G4 motifs in RKO control cells (Fig. 5d). Next, we monitored G4 density in the domains where replication timing had been changed following Pol θ depletion. As G4 motifs strongly correlate with replication origins, measuring G4 density allows an indirect assessment of the number of potential replication origins in E, EtoL, LtoE and L domains. As for Line-1 content, gene density and GC content, EtoL and LtoE domains showed an intermediate G4 content (Fig. 5d), reflecting an intermediate level of potential origins. The characteristics of all the domains EtoL and LtoE are presented in Supplementary Tables 1 and 2, respectively.

These analyses strengthen the notion that regions where replication timing can change possess specific genomic signatures, as Gilbert and colleagues³¹ have previously proposed, and further support the model that Pol θ is a regulator of replication timing of particular regions of the human genome.

Excess Pol θ delays replication timing. To obtain more insight into the mechanism by which Pol θ contributes to replication timing, we analysed the impact of the excess of Pol θ on replication timing at a genome-wide level. We generated human cells that stably overexpress Pol θ compared with the endogenous Pol θ level of the isogenic control²³ (Fig. 6a) and examined replication timing as described above. Excess Pol θ resulted mostly in a delayed timing of multiple domains (a global analysis is shown in Supplementary Fig. 7; two examples are presented in Fig. 6b). It is important to note that as seen with Pol θ depletion, overexpression of Pol θ had no impact on the global level of Rif1 (Fig. 6a). These results clearly show that excess Pol θ delays the replication timing of a subset of domains.

Discussion

Shortly after mitosis, daughter cells organize their genomes into large replication domains containing multiple replication initiation sites that are activated simultaneously. The organization of replication domains occurs during the G1 phase of the cell cycle at the TDP. Although orderly progression of DNA replication is important to ensure stable genetic and epigenetic inheritance, the mechanisms underlying the timing programme of replication have yet to be fully elucidated. Here we have identified a novel actor of the human DNA replication timing programme, Pol θ , an enzyme classified thus far as a specialized repair A-family DNA polymerase^{2,42}. We first presented evidence that the timing of association of Pol θ with chromatin appears to be similar to that of pre-RC formation and is concomitant with Mcm loading. Moreover, we show that Pol θ interacts with Orc2 and Orc4, subunits of the Orc platform for the loading of pre-RC components. Finally, Pol θ -depleted cells demonstrated multiple abnormalities: increased Mcm loading onto chromatin in G1 and genome-wide changes in replication timing of a subset of domains with no modification of origin density. All these data are consistent with the hypothesis that Pol θ plays a role at or near the TDP by regulating the number of Mcm molecules at individual pre-RCs. Interestingly, computational analysis of replication timing has previously indicated that replication timing can be regulated by the number of Mcm molecules bound to the origin and the chromatin context⁴³. This model predicts that origins with more loaded Mcm have a higher probability of firing and thus an earlier average replication time⁴³. We propose the following model to explain our results: when Pol θ is expressed, it could bind to the Orc at a subset of individual late origins in G1 and inhibits the recruitment of excess Mcm to avoid early replication timing of these late origins. When Pol θ is depleted, increased Mcm loading at these individual late origins

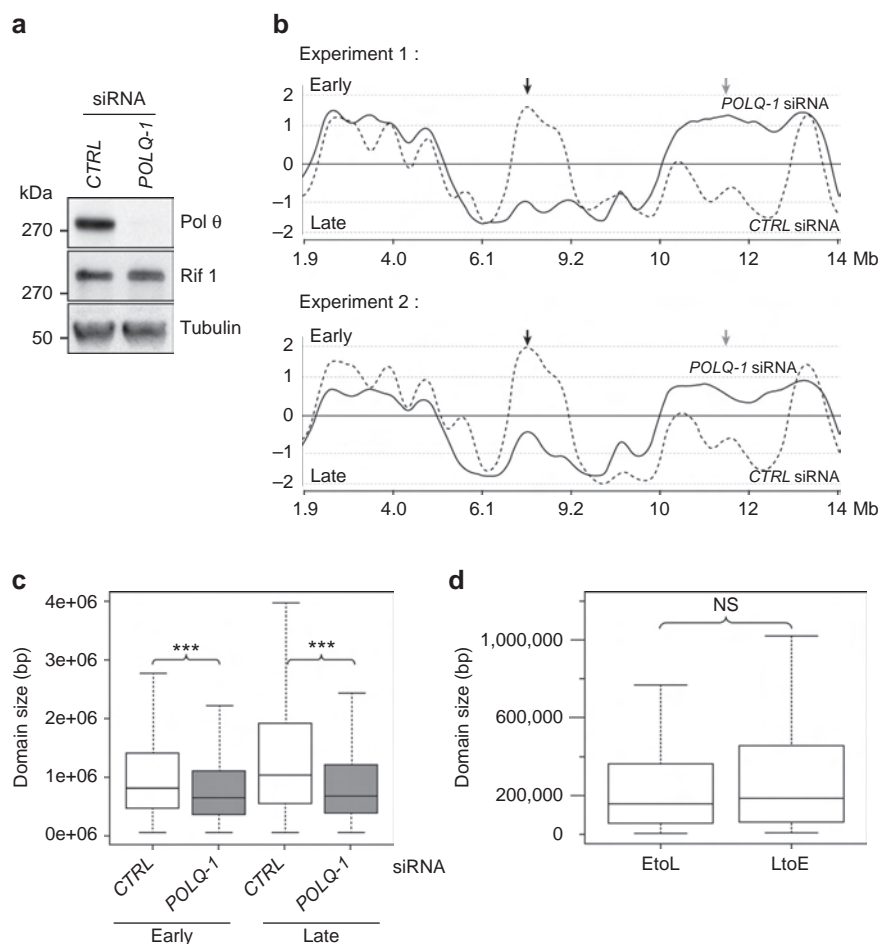


Figure 4 | Pol θ depletion modifies the temporal replication programme. (a) Western blot analysis with the indicated antibodies of whole-cell extracts from mock- (CTRL siRNA) and Pol θ - (POLQ-1 siRNA) depleted RKO cells used for the replication timing experiments shown in b and Supplementary Fig. 2. (b) The replication timing was analysed 24 h after transfection with CTRL or POLQ-1 siRNA. Shown are the microarray profiles of the timing of replication on chromosome 6: 1 997 799–14 527 487 (12.5 Mb, hg18) from two independent experiments. Black lines represent POLQ-1 siRNA and dotted lines represent control (CTRL) siRNA. Arrows indicate examples of changes in replication timing. Black arrows: early to late (EtoL); grey arrows: LATE TO Early (LtoE). The size of the replication domains (c) and the size of the domains where replication timing was changed (d) were analysed. The thick line represents the median, the limit of the box corresponds to the 0.25 to 0.75 quartiles with whiskers extending to the maximum value of 1.5 times the interquartile range; Early: early domains, Late: late domains, EtoL: early-to-late, LtoE: late-to-early; *** P -value < 0.001. Data are representative of two independent experiments of microarrays for each siRNA experiments and thus generate more than 700 domains for each categories. Statistical significance was performed with Student's t -test from the R package. The characteristic of the t -test in c is: $t = 6.0954$, $df = 1750.042$, P -value = $1.34e - 09$, indicating here the significance of the test. Uncropped blots are shown in Supplementary Fig. 8.

could be facilitated, possibly via enhanced interaction with Orc, resulting in the observed LtoE switch. The abnormal recruitment of activating limiting factors to these advanced late origins could reduce the full availability of these factors locally at some early origins, explaining why the activation of some early origins are delayed (EtoL). When Pol θ is overexpressed, it could bind also early-middle origins through Orc interaction and consequently delay their activation. We therefore propose that in human cells Pol θ establishes some late S-phase replication domains that are restrained from being activated at early S phase.

The discovery that Pol θ contributes to the regulation of DNA replication timing in human cells is of particular interest, as the regulation of this programme is extremely robust. In vertebrates, only the loss of Rif1 has been shown thus far to affect replication timing of a subset of domains^{32,34}. Rif1 was originally discovered as a telomere-binding protein in yeast⁴⁴ and a double-strand break response factor in mammals⁴⁵, and has previously been shown to regulate the timing of replication in yeast³⁵. Importantly, our data suggest that although Pol θ and Rif1

display similar binding kinetics to chromatin at TDP when cells exit mitosis, their roles in controlling replication timing seem to be different. Indeed, in contrast to Pol θ , Rif1 seems to establish mid S-phase replication domains that are restrained from being activated at early S phase, without any impact on pre-RC formation^{32,34}. Moreover, Pol θ depletion affects neither the global level of Rif1 nor its recruitment to chromatin in G1.

Pol θ is encoded by the *POLQ* gene that is present in plants and multicellular eukaryotes but not yeast or other fungi⁴⁶. Although mammalian Pol θ is often best known for its proposed role in repair of genome damage, it appears that there are important functional differences between species and cell types. Moreover, very little of its physiological function in unstressed cells has been thoroughly investigated. The *TEBICHI* (*TEB*) gene, encoding a Pol θ -like protein found in *Arabidopsis thaliana*, is required for normal plant development in the absence of external stress^{47,48}. Both the *POLQ* knockout mouse⁴⁹ and the chaos1 (chromosome aberration occurring spontaneously 1) mouse that possesses a mutant allele of *POLQ*⁵⁰ show increased manifestations of

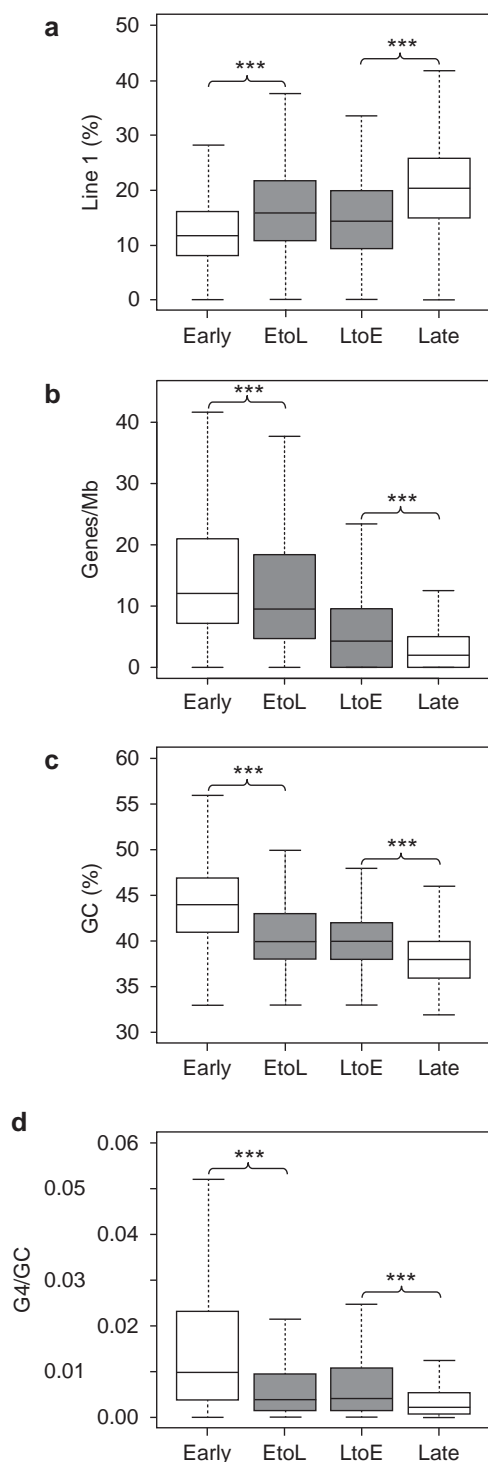


Figure 5 | Characterization of the domains in which replication timing was changed following Pol θ depletion. Analysis of Line-1 content (**a**), gene density (**b**), GC content (**c**) and G4/GC content (**d**) in Early domains, early-to-late domains (EtoL), late-to-early domains (LtoE) and late domains in Pol θ -depleted cells. The statistical significance was performed with Student's *t*-test from the R package. The thick line represents the median, the limit of the boxes corresponds to the 0.25 to 0.75 quartiles with whiskers extending to the maximum value of 1.5 times the interquartile range. ****P*-value < 0.001. Data are representative of two independent microarrays experiments for each siRNA, generating more than 700 domains for each category. The *t*-test characteristics are $t = 3.3644$, $df = 298.512$, $P\text{-value} = 8.6e - 4$ for **a**; $t = 6.78$, $df = 314.213$, $P\text{-value} = 5.9e - 17$ for **b**; $t = 5.0732$, $df = 557.542$, $P\text{-value} = 5.335e - 07$ for **c**; and $t = 5.0732$, $df = 557.542$, $P\text{-value} = 5.335e - 07$ for **d**, indicating the significance of the test.

spontaneous genetic instability, with enhanced formation of micronuclei. Interestingly, in the same screen that identified the *chaos1* allele of *POLQ*, a second mutant called *chaos3* exhibited a similar phenotype of increased micronuclei⁵¹. *Chaos3* was identified as a hypomorphic allele of Mcm4, a component of the hexameric Mcm helicase complex.

Our data also provide new information on the role of Pol θ during the cell cycle in unstressed cells, and extend the function of Pol θ to include normal DNA replication, strengthening the concept that TLS or repair synthesis after external stress might not be the sole function assigned to non-replicative DNA polymerases. We recently showed that this is the case for both Pol η , which is required for the stability of common fragile sites during unperturbed S-phase^{52,53}, and Pol κ , which functions at stalled forks in unstressed cells to ensure high cell viability and genomic stability⁵⁴.

This novel Pol θ function may be particularly important within the context of cancer, as high *POLQ* expression is strongly associated with poor clinical outcome of patients with colon, breast and lung cancers^{22–24,55,56}. Abnormal temporal control of replication, recently demonstrated in several haematologic malignancies, has been proposed to represent a common mechanism responsible for genomic instability found in cancer cells⁵⁷; therefore, Pol θ overexpression may be of cardinal importance in defining the modified temporal replication programme in cancer cells and may constitute a driving force for chromosomal instability^{58–60}.

Methods

Cell line and culture conditions. The human RKO cell line was purchased from ATCC and grown in DMEM with GlutaMAX I, high-glucose, sodium pyruvate (Gibco, Life Technologies, Paisley, UK), supplemented with 10% fetal bovine serum (Lonza, Basel, CH), penicillin (100 U ml⁻¹) and streptomycin (100 μ g ml⁻¹) (Gibco) at 37 °C, 5% CO₂ and 5% O₂ (standard culture conditions). The design and construction of the *POLQ* expression vector as well as the overexpressing Pol θ cells (Q3) and their isogenic controls (Ctrl2) were previously described²³. They were grown in Modified Eagle's Medium- α with GlutaMAX I, high-glucose, sodium pyruvate (Gibco, Life Technologies) supplemented with 10% fetal bovine serum (Lonza), hygromycin B (50 μ g ml⁻¹), penicillin (100 U ml⁻¹) and streptomycin (100 μ g ml⁻¹) (Gibco) at 37 °C, 5% CO₂ and 5% O₂.

Transfection of siRNA. RKO cells were transfected with siRNA targeting either *POLQ* or *Luciferase* (*CTRL*). *POLQ-1* is an siRNA ON-TARGET plus smart pool of four *POLQ* siRNA from Dharmacon (Thermo Fisher Scientific, Waltham, USA): 5'-CCUUAAGACUGUAGGUACU-3', 5'-ACACAGUAGGCGAGAGU-3', 5'-CGACUAAAGAUAGAUAUUU-3' and 5'-CAAACAACCCUUAUCGUAA-3'. *POLQ-2* is an siRNA ON-TARGET plus smart pool of four *POLQ* siRNA from Dharmacon: 5'-UCAGAGGGAUGGAGCUAAU-3', 5'-GAGAUUACCCUUUCACCUA-3', 5'-AAUUUUACAGUACGGAAA-3' and 5'-UGAUAGAUUAGC CUAGUA-3'. For *Luciferase* (*CTRL*), 5'-CGUACGCGGAUACUUCGAdTdT-3' (Sigma-Aldrich, Saint-Louis, USA) was used. siRNAs were used at a final concentration of 50 nM, using the Lipofectamine 2000 transfection reagent (Invitrogen, Carlsbad, USA) following the manufacturer's suggestions. The transfection medium was replaced 8 h later by complete medium.

Cell synchronization. Cells were synchronized in mitosis by nocodazole (0.1 μ M, Sigma-Aldrich) for 20 h. Next, cells were collected by mitotic shake-off, washed twice in warm complete medium and released into the cell cycle under standard culture conditions. At the indicated times, cells were collected for subcellular fractionation and BrdU incorporation assay. When cells were transfected with *POLQ-1* or *CTRL* siRNA, nocodazole was added to the medium 30 h later. For cell synchronization in G2/M with RO-3306, cells were incubated for 20 h with RO-3306 (9 μ M, Sigma-Aldrich), then released after 2 washes in warm complete medium. Thirty minutes later, mitotic cells were collected by mitotic shake-off and released into the cell cycle under standard culture conditions. At the indicated times, cells were collected for subcellular fractionation and a BrdU incorporation assay. As a control, asynchronous cells were incubated with dimethyl sulphoxide (Sigma-Aldrich) instead of RO-3306. When cells were transfected with *POLQ-2* or *CTRL* siRNA, RO-3306 was added 26 h after transfection.

BrdU incorporation assay. Cells were pulse-labelled with 10 μ M BrdU (Sigma-Aldrich) for 15 min then collected by trypsinization and fixed in 70% ethanol

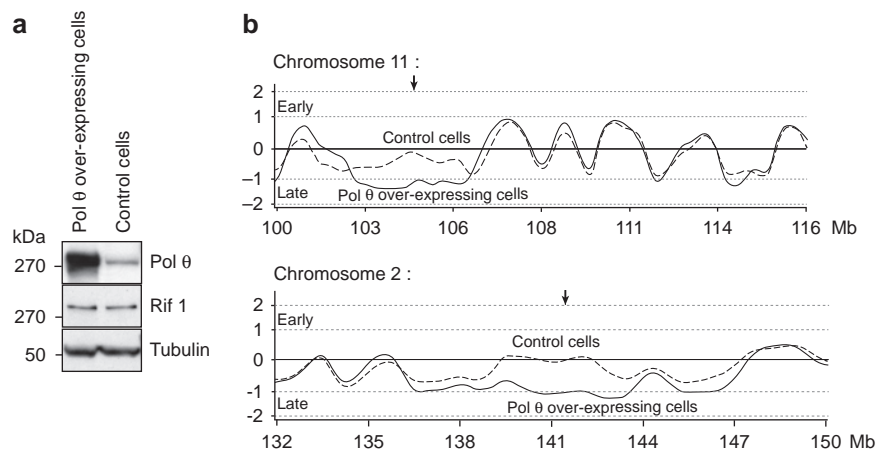


Figure 6 | Excess Pol θ leads to a delayed replication of several replicating domains. (a) Western blot analysis with the indicated antibodies of whole-cell extracts from control and overexpressing MRC5 cells used for the replication timing experiments shown in b and Supplementary Fig. 4. (b) Two examples of microarray profiles of the timing of replication on chromosomes 11 and 2 in Pol θ -overexpressing cells and isogenic control cells. Uncropped blots are shown in Supplementary Fig. 8.

overnight at -20°C . After washing with PBS-BSA 0.5% (Sigma-Aldrich), DNA was denatured for 20 min with 2 N HCl treatment (Normapur, VWR, Lutterworth, UK) at room temperature (RT). Cells were washed in PBS-BSA 0.5%, incubated with sodium tetraborate (0.1 M, pH 8.5) for 2 min at RT, washed and then incubated with a mouse anti-BrdU fluorescein isothiocyanate antibody (1:25, Becton Dickinson) in PBS-BSA 0.5%, Tween20 0.5%, for 1 h at RT. After washing, Alexa Fluor 488 goat anti-mouse IgG (1:50, Invitrogen) was added in PBS-BSA 0.5%, Tween20 0.5% for 30 min at RT. Finally, after washing in PBS-BSA 0.5%, then in PBS, cells were resuspended in PBS with propidium iodide ($25\text{ }\mu\text{g ml}^{-1}$, Sigma-Aldrich) and RNase A ($100\text{ }\mu\text{g ml}^{-1}$, Thermo Fisher Scientific). After 20 min of incubation at RT, cell cycle analysis was carried out by flow cytometry with a FACScalibur cytometer (Becton Dickinson).

Whole-cell extracts. Cells were resuspended in lysis buffer (Tris-HCl: 50 mM, pH 7.5, NaCl 250 mM, EDTA 1 mM, Triton X-100 0.1%, with protease inhibitors) for 30 min on ice. After centrifugation at 15,000 rcf, for 20 min at 4°C , the supernatant was collected. Laemmli buffer was added and the mixture boiled for 3 min.

Cell fractionation. Subcellular fractionation was performed as described previously⁶¹. Briefly, cells were lysed in Buffer A (Hepes 10 mM pH 7.9, KCl 10 mM, MgCl_2 1.5 mM, sucrose 0.34 M, Glycerol 10%, dithiothreitol (DTT) 1 mM, Halt protease and phosphatase inhibitor cocktail from Thermo Fisher Scientific) complemented with Triton X-100 0.1% for 5 min on ice. After centrifugation at 1,500 rcf, 5 min, 4°C , the supernatant was clarified by high-speed centrifugation (18,000 rcf, 4°C , 15 min) to obtain the cytoplasmic fraction. The pellet was washed once with Buffer A and then incubated in Buffer B (EDTA 3 mM, EGTA 0.2 mM, DTT 1 mM, Halt protease and phosphatase inhibitor cocktail from Thermo Fisher Scientific) for 30 min on ice. After centrifugation (1,700 rcf, 5 min, 4°C), the supernatant was collected as the soluble nuclear fraction. The pellet was washed once with Buffer B and resuspended in the same buffer, then sonicated three times for 10 s at 25% amplitude with a Sonics Vibra Cell Ultrasonic processor. This fraction included the chromatin-enriched fraction. Laemmli buffer was added and cell extracts were boiled for 3 min.

Western blot analysis and immunoprecipitation. Proteins were separated in 8% Tris-Glycine, NuPAGE Novex 3–8% Tris-Acetate, or bolt 4–12% bis tris gels (Invitrogen) according to supplier instructions and then transferred onto polyvinylidene difluoride membranes (GE Healthcare, Fairfield, USA) in CAPS transfer buffer. For subcellular fractionation, equivalent cell number fraction samples were loaded. Saturation of the membrane was achieved by soaking for 1 h in PBS Tween 0.1% containing 5% non-fat milk. The antibodies used for immunodetection were: Actin (MAB 1501, 1:50,000, Merck Millipore, Darmstadt, DE), Cdc6-180.2 (sc-9964, 1:500, Santa Cruz Biotechnology, Dallas, USA), Cdt1 (8064, 1:1,000, Cell Signaling, Danvers, USA), Mcm2 (ab4461, 1:2,000, Abcam), Mcm4-G7 (sc-28317, 1:2,000, Santa Cruz Biotechnology), Mcm5 (ab17967, 1:5,000, Abcam), Mcm7 (sc-22782, 1:1,000, Santa Cruz Biotechnology), Orc2 (M055-3, 1:1,000, MBL, Woburn, USA, or 4736, 1:1,000, Cell Signaling), Orc4 (– 2: Santa Cruz-Sc-136331, 1:1,000), Rif1 (A300–569A, 1:1,000, Bethyl, Montgomery, USA), Pol Alpha (ab311777, 1:250, Abcam), Pol Delta (sc-17776, 1:500, Santa Cruz Biotechnology), Tubulin B-5-1-2 (T6074, Sigma Aldrich), DNA-PKcs (ab1832, 1:5,000, Abcam, Cambridge, USA), Chk1 (sc-8408, 1:1,000, Santa Cruz Biotechnology), Phospho-Chk1 (Ser345) (2348, 1:1,000, Cell Signaling), RPA (NA18, 1:1,000, Calbiochem),

Phospho-RPA32 (S4/S8) (A300-245A, 1:1,000, Bethyl Laboratories), Chk2 (sc-17748, 1:1,000, Santa Cruz Biotechnology), Phospho-Chk2 (Thr68) (2197, 1:1,000, Cell Signaling), horseradish peroxidase (HRP)-conjugated anti-rabbit and HRP-conjugated anti-mouse (1:10,000, Jackson Immuno Research, West Grove, USA), HRP-conjugated anti-rat (1:5,000). Pol θ -antibodies Pol θ -1 (western blot analysis) and Pol θ -2 (immunoprecipitation experiment) are mouse monoclonal antibodies, affinity purified with protein A ceramic hyperDF (Akta System) ($0.1\text{ }\mu\text{g ml}^{-1}$). They were raised against the peptide antigens CSIFRARKRASLDINKEKPG and HDETSSLLPRKESNIVDDNGC, respectively, derived from regions of the central domain of POLQ. Their specificity was checked using a siRNA strategy.

For immunoprecipitation experiments, 20×10^6 cells were lysed in 1 ml of IP buffer (Tris 50 mM pH8, NaCl 150 mM, EGTA 3 mM, NP40 1%, protease and phosphatase inhibitors, and kept on ice for 30 min. After centrifugation at 13,000 r.p.m. for 10 min at 4°C , the supernatant was pre-cleared with 40 μl protein A and G (or G only) dynabeads (Invitrogen) for 1 h at 4°C and then incubated with 8 μg of anti-Orc2-1 (Abcam—ab68348), 5 μg of anti-Orc4 (1: Tranduction Laboratories—H83120, 1:1,000; 2: Santa Cruz-Sc-136331, 1:1,000) or Pol θ -2 antibodies for 150 min and 40 μl of protein A and G (or only G) dynabeads (Invitrogen) was added for two additional hours. Following three washes with the IP buffer, the immunoprecipitated material was resuspended in 60 μl of Laemmli sample buffer, boiled for 3 min and analysed by western blotting. For experiments revealing Pol θ and Orc2 interaction, Pol θ immunoprecipitation was performed as followed: RKO cell pellets were incubated in hypotonic buffer (10 mM Hepes pH 7.9, 1.5 mM MgCl_2 , 10 mM KCl, 0.5 mM DTT, phenylmethylsulphonyl fluoride (PMSF) 0.2 mM) and pelleted by centrifugation at 4,600 r.p.m. for 5 min at 4°C . They were then resuspended in hypotonic buffer on ice 10 min, homogenized in a glass dounce homogenizer and nuclei were collected by centrifugation at 6,100 r.p.m. for 15 min at 4°C . Nuclei (volume v) were then resuspended in v/2 low-salt buffer (20 mM Hepes pH 7.9, 25% glycerol, 1.5 mM MgCl_2 , 20 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, PMSF 0.2 mM) and v/2 high-salt buffer (20 mM Hepes pH 7.9, 25% glycerol, 1.5 mM MgCl_2 , 1.2 M KCl, 0.2 mM EDTA, 0.5 mM DTT, PMSF 0.2 mM) was added. They were then extracted for 30 min and pelleted at 14,500 r.p.m. for 25 min at 4°C . The supernatant (nuclear extract) was then dialysed in dialysis buffer (20 mM Hepes pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, PMSF 0.2 mM) for 90 min at 4°C . Finally, the dialysed nuclear extract was clarified by centrifugation at 14,500 r.p.m. for 20 min at 4°C , and benzonase (25 U) was then added.

Estimation of origin activity by DNA combing. Twenty-four hours after transfection of the RKO cell line with control (CTRL) or POLQ (POLQ-1) siRNA, RKO cells were labelled successively with 50 μM IdU (Sigma-Aldrich) for 15 min, 100 μM CldU for 15 min and 3 h 200 μM thymidine. Genomic DNA was prepared in agarose plugs (0.5×10^5 cells per plug) and DNA combing was performed as described previously^{62,63}, except that agarose plugs were melted for 30 min at 65°C before adding 8 U of β -agarase (New England Biolabs) for overnight digestion at 42°C . IdU, CldU and single-stranded DNA were detected with mouse anti-BrdU (347580, 1:20, Becton Dickinson) and Alexa 546 Goat anti-Mouse (A11030, 1:50, Molecular Probes), Rat anti-BrdU (Abc117–7513, 1:20, Seralab) and Alexa 488 Goat anti-Rat IgG (A11017, 1:50, Molecular Probes), and Mouse anti-ssDNA (MAB3034, 1:300, Millipore) and Alexa 647 Goat anti-Mouse IgG2a (A21241, 1:50, Molecular Probes), respectively. IODs and Global Instant Fork density were measured as described^{63,64}. Seventeen and 12 IOD, as well as 126 and 89 Forks were measured for CTRL and Pol θ siRNA cells, respectively. The non-parametric

test of Mann–Whitney was applied to compare data sets from control and Pol θ -depleted cells with the use of 12 forks subgroups for the fork density analysis.

Neo-synthesized DNA preparation and microarrays. RKO cells transfected with *POLQ-1* or *CTRL* siRNA for 24 h, and MRC5-SV cells that stably overexpress Pol θ and its isogenic control cells, were incubated with BrdU (50 μ M) for 1 h then collected, washed three times in PBS, fixed in 75% ethanol and stored at -20°C . Fixed cells were re-suspended first in PBS with RNase (0.5 mg ml $^{-1}$) and then with propidium iodide (50 μ g ml $^{-1}$) followed by an incubation of 30 min at RT. Cells (80,000) were sorted in two fractions, S1 and S2, using INFLUX 500 (Cytospea BD Biosciences) corresponding to early and late S-phase fractions, respectively. Neo-synthesized DNA was immunoprecipitated with BrdU antibodies (Anti-BrdU Pure, BD Biosciences, 347580) as described⁶⁵. To control the quality of enrichment of early and late fractions in S1 and S2, quantitative PCR was performed with CAV2 oligonucleotides (early control) and with bgGRM8 oligonucleotides (late control). Microarray hybridization requires a minimum of 500 ng of DNA and to obtain sufficient specific immunoprecipitated DNA, whole-genome amplification was conducted (WGA, Sigma). To be sure that this step did not introduce bias, an after WGA quantitative PCR was performed to confirm the specific enrichment in both the S1 and S2 fractions. After amplification, early and late neo-synthesized DNA were labelled with Cy3 and Cy5 ULS molecules (Genomic DNA Labelling Kit, Agilent) as recommended by the manufacturer. The hybridization was performed according to the manufacturer's instructions on 4 \times 180 K human microarrays (SurePrint G3 Human CGH Microarray Kit, 4 \times 180 K, Agilent Technologies, genome reference Hg18) that covers the whole genome with one probe every 13 Kb (11 Kb in RefSeq sequences). Microarrays were scanned with an Agilent High-Resolution C Scanner using a resolution of 2 μ m and the autofocus option. Feature extraction was performed with the Feature Extraction 9.1 software (Agilent Technologies). For each experiment, the raw data sets were automatically normalized by the Feature extraction software. Analysis was performed with the Agilent Genomic Workbench 5.0 software. The log $_2$ -ratio timing profiles were smoothed using the Triangular Moving Average option of the Agilent Genomic Workbench 5.0 software with the linear algorithm and 500-kb windows.

Identification of replication domains. To determine the replication domains in Pol θ -depleted RKO cells and in Pol θ -overexpressing MRC5-SV cells, the algorithms from CGH applications in the Agilent Genomic Workbench 5.0 software were used, particularly the Aberration detection algorithms that define the boundaries and magnitudes of the regions of DNA loss or gain when integrating the Z-score. It finds genomic intervals with an over- or underabundance of probes with log ratios that deviate significantly from baseline with a threshold of 1.8. Results from the Z-score identify regions of enriched probe log ratio change, and can suggest early domains when the log-ratio is positive and late domain when the log-ratio is negative. A comparison was conducted between early and late domains from both cell lines to determine segments where replication timing changes. As mentioned by Hiratani *et al.*³¹, two major changes can be characterized. The intersection between *CTRL* siRNA late and *POLQ-1* siRNA early intervals will reveal segments that have become earlier by the depletion of Pol θ , and conversely with *CTRL* siRNA early versus *POLQ-1* siRNA late domains, intervals that have become later will be characterized. Similarly, the intersection between MRC5-*CTRL* late and MRC5-overexpressing Pol θ early intervals will reveal segments that have become earlier by the overexpression of Pol θ , and conversely with MRC5-*CTRL* early versus MRC5-overexpressing Pol θ late domains, intervals that have become later will be characterized. The intersection was performed with GALAXY tools (<https://main.g2.bx.psu.edu/>)⁶⁶.

Line-1 and GC contents. The Line-1 element content was extracted with the Profile Annotation tool from GALAXY. For the GC content, two steps were required: (i) determination of the DNA sequence from intervals performed with the Extract Genomic DNA tool; and (ii) calculation of the GC content by the GeeCee EMBOS tool installed on the GALAXY website. For the Line-1 content, we used the coverage tool from the GALAXY website.

G4/GC content. The coordinates of the G4 in the human genome (hg 18) were determined by Python script as described in: <http://code.google.com/p/bioinformatics-misc/source/browse/trunk/quadparser.py?spec=svn3&r=3>. Next, G4 content was obtained with the coverage tool and normalization with the GC content was performed with the result obtained above.

References

- Hogg, M., Sauer-Eriksson, A. E. & Johansson, E. Promiscuous DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* **40**, 2611–2622 (2012).
- Seki, M., Marini, F. & Wood, R. D. POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* **31**, 6117–6126 (2003).
- Prasad, R. *et al.* Human DNA polymerase theta possesses 5'-dRP lyase activity and functions in single-nucleotide base excision repair in vitro. *Nucleic Acids Res.* **37**, 1868–1877 (2009).
- Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B. & Kunkel, T. A. Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* **36**, 3847–3856 (2008).
- Seki, M., Gearhart, P. J. & Wood, R. D. DNA polymerases and somatic hypermutation of immunoglobulin genes. *EMBO Rep.* **6**, 1143–1148 (2005).
- Maga, G., Shevlev, I., Ramadan, K., Spadari, S. & Hubscher, U. DNA polymerase theta purified from human cells is a high-fidelity enzyme. *J. Mol. Biol.* **319**, 359–369 (2002).
- Masuda, K. *et al.* Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. *DNA Repair (Amst)* **5**, 1384–1391 (2006).
- Masuda, K. *et al.* DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes. *Proc. Natl Acad. Sci. USA* **102**, 13986–13991 (2005).
- Zan, H. *et al.* The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. *EMBO J.* **24**, 3757–3769 (2005).
- Martomo, S. A., Saribasak, H., Yokoi, M., Hanaoka, F. & Gearhart, P. J. Reevaluation of the role of DNA polymerase theta in somatic hypermutation of immunoglobulin genes. *DNA Repair (Amst)* **7**, 1603–1608 (2008).
- Yoshimura, M. *et al.* Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. *Mol. Cell* **24**, 115–125 (2006).
- Asagoshi, K. *et al.* Single-nucleotide base excision repair DNA polymerase activity in *C. elegans* in the absence of DNA polymerase beta. *Nucleic Acids Res.* **40**, 670–681 (2012).
- Goff, J. P. *et al.* Lack of DNA polymerase theta (POLQ) radiosensitizes bone marrow stromal cells in vitro and increases reticulocyte micronuclei after total-body irradiation. *Radiat. Res.* **172**, 165–174 (2009).
- Higgins, G. S. *et al.* A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res.* **70**, 2984–2993 (2010).
- Ukai, A. *et al.* Role of DNA polymerase theta in tolerance of endogenous and exogenous DNA damage in mouse B cells. *Genes Cells* **11**, 111–121 (2006).
- Chan, S. H., Yu, A. M. & McVey, M. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* **6**, e1001005 (2010).
- Yu, A. M. & McVey, M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res.* **38**, 5706–5717 (2010).
- Hogg, M., Seki, M., Wood, R. D., Double, S. & Wallace, S. S. Lesion bypass activity of DNA polymerase theta (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. *J. Mol. Biol.* **405**, 642–652 (2011).
- Seki, M. *et al.* High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J.* **23**, 4484–4494 (2004).
- Masuda, K. *et al.* DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J. Biol. Chem.* **282**, 17387–17394 (2007).
- Seki, M. & Wood, R. D. DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6-4) photoproduct. *DNA Repair (Amst)* **7**, 119–127 (2008).
- Allera-Moreau, C. *et al.* DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis* **1**, e30 (2012).
- Lemee, F. *et al.* DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl Acad. Sci. USA* **107**, 13390–13395 (2010).
- Pillaire, M. J. *et al.* A 'DNA replication' signature of progression and negative outcome in colorectal cancer. *Oncogene* **29**, 876–887 (2010).
- Hoffmann, J. S. & Cazaux, C. Aberrant expression of alternative DNA polymerases: a source of mutator phenotype as well as replicative stress in cancer. *Semin. Cancer Biol.* **20**, 312–319 (2010).
- Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
- Wu, P. Y. & Nurse, P. Establishing the program of origin firing during S phase in fission yeast. *Cell* **136**, 852–864 (2009).
- Clijsters, L., Ogink, J. & Wolthuis, R. The spindle checkpoint, APC/CCdc20, and APC/CCdh1 play distinct roles in connecting mitosis to S phase. *J. Cell Biol.* **201**, 1013–1026 (2013).
- Pillaire, M. J. *et al.* Upregulation of error-prone DNA polymerases beta and kappa slows down fork progression without activating the replication checkpoint. *Cell Cycle* **6**, 471–477 (2007).
- Gilbert, D. M. *et al.* Space and time in the nucleus: developmental control of replication timing and chromosome architecture. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 143–153 (2010).
- Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).

32. Cornacchia, D. *et al.* Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *EMBO J.* **31**, 3678–3690 (2012).
33. Schmegner, C., Hameister, H., Vogel, W. & Assum, G. Isochores and replication time zones: a perfect match. *Cytogenet. Genome Res.* **116**, 167–172 (2007).
34. Yamazaki, S. *et al.* Rif1 regulates the replication timing domains on the human genome. *EMBO J.* **31**, 3667–3677 (2012).
35. Hayano, M. *et al.* Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes Dev.* **26**, 137–150 (2012).
36. Hassan-Zadeh, V. *et al.* USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS Biol.* **10**, e1001277 (2012).
37. Guilbaud, G. *et al.* Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput. Biol.* **7**, e1002322 (2011).
38. Bernardi, G. Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3–17 (2000).
39. Costantini, M. & Bernardi, G. Replication timing, chromosomal bands, and isochores. *Proc. Natl Acad. Sci. USA* **105**, 3433–3437 (2008).
40. Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
41. Cayrou, C. *et al.* New insights into replication origin characteristics in metazoans. *Cell Cycle* **11**, 658–667 (2012).
42. Sharief, F. S., Vojta, P. J., Ropp, P. A. & Copeland, W. C. Cloning and chromosomal mapping of the human DNA polymerase theta (POLQ), the eighth human DNA polymerase. *Genomics* **59**, 90–96 (1999).
43. Yang, S. C., Rhind, N. & Bechhoefer, J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Mol. Syst. Biol.* **6**, 404 (2010).
44. Hardy, C. F., Sussel, L. & Shore, D. A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev.* **6**, 801–814 (1992).
45. Silverman, J., Takai, H., Buonomo, S. B., Eisenhaber, F. & de Lange, T. Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint. *Genes Dev.* **18**, 2108–2119 (2004).
46. Yousefzadeh, M. J. & Wood, R. D. DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair (Amst)* **12**, 1–9 (2013).
47. Inagaki, S., Nakamura, K. & Morikami, A. A link among DNA replication, recombination, and gene expression revealed by genetic and genomic analysis of TEB1CHI gene of Arabidopsis thaliana. *PLoS Genet.* **5**, e1000613 (2009).
48. Inagaki, S. *et al.* Arabidopsis TEB1CHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell.* **18**, 879–892 (2006).
49. Shima, N., Munroe, R. J. & Schimenti, J. C. The mouse genomic instability mutation chaos1 is an allele of Polq that exhibits genetic interaction with Atm. *Mol. Cell Biol.* **24**, 10381–10389 (2004).
50. Shima, N. *et al.* Phenotype-based identification of mouse chromosome instability mutants. *Genetics* **163**, 1031–1040 (2003).
51. Shima, N. *et al.* A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat. Genet.* **39**, 93–98 (2007).
52. Bergoglio, V. *et al.* DNA synthesis by Pol eta promotes fragile site stability by preventing under-replicated DNA in mitosis. *J. Cell Biol.* **201**, 395–408 (2013).
53. Rey, L. *et al.* Human DNA polymerase eta is required for common fragile site stability during unperturbed DNA replication. *Mol. Cell Biol.* **29**, 3344–3354 (2009).
54. Betous, R. *et al.* DNA polymerase kappa-dependent DNA synthesis at stalled replication forks is important for CHK1 activation. *EMBO J.* **32**, 2172–2185 (2013).
55. Higgins, G. S. *et al.* Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* **1**, 175–184 (2010).
56. Kawamura, K. *et al.* DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *International journal of cancer. J. Int. Cancer* **109**, 9–16 (2004).
57. Ryba, T. *et al.* Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res.* **22**, 1833–1844 (2012).
58. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
59. Durkin, S. G. & Glover, T. W. Chromosome fragile sites. *Annu. Rev. Genet.* **41**, 169–192 (2007).
60. Letessier, A. *et al.* Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120–123 (2011).
61. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell Biol.* **20**, 8602–8612 (2000).
62. Michalet, X. *et al.* Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **277**, 1518–1523 (1997).
63. Schwob, E. *et al.* Use of DNA combing for studying DNA replication in vivo in yeast and mammalian cells. *Methods. Mol. Biol.* **521**, 673–687 (2009).
64. Guillou, E. *et al.* Cohesin organizes chromatin loops at DNA replication factories. *Genes Dev.* **24**, 2812–2822 (2010).
65. Prioleau, M. N., Gendron, M. C. & Hyrien, O. Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. *Mol. Cell Biol.* **23**, 3536–3549 (2003).
66. Giardine, B. *et al.* Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451–1455 (2005).

Acknowledgements

We gratefully acknowledge Marcel Méchali (IGH—Montpellier), leader of the French 'Replication of eukaryotic chromosomes and its checkpoint' GDR 2915—CNRS Network, which allowed the adequate collaborations to develop this project. We thank Marie-Jeanne Pillaire, Christine Bordier, Renaud Poincloux, Françoise Viala (IPBS—Toulouse) and Nicole Bogetto (IJM—Paris) for their technical support. We also acknowledge Kelly Thornber (www.scientificscripts.com) and Anne-Lise Haenni (IJM—Paris) for critical reading of the manuscript. This work was supported by La Ligue contre le cancer (Equipe Labellisée 2013 to J.-S.H.; Comité Ile-de-France to G.B.) and a GEFLUC grant (to J.-C.C.).

Author contributions

A.F.-V., J.-C.C., C.C. and J.-S.H. designed the experiments. A.F.-V., L.G.S., J.-C.C. and M.D. performed the experiments. A.F.-V., J.-C.C., E.S., G.B., C.C. and J.-S.H. analysed the data and contributed to the manuscript preparation. J.-S.H. supervised the project and wrote the manuscript.

Additional information

Accession codes: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo (accession no. GSE49693 for depleted Polθ experiments and accession number GSE53070 for overexpressed Pol θ experiments).

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

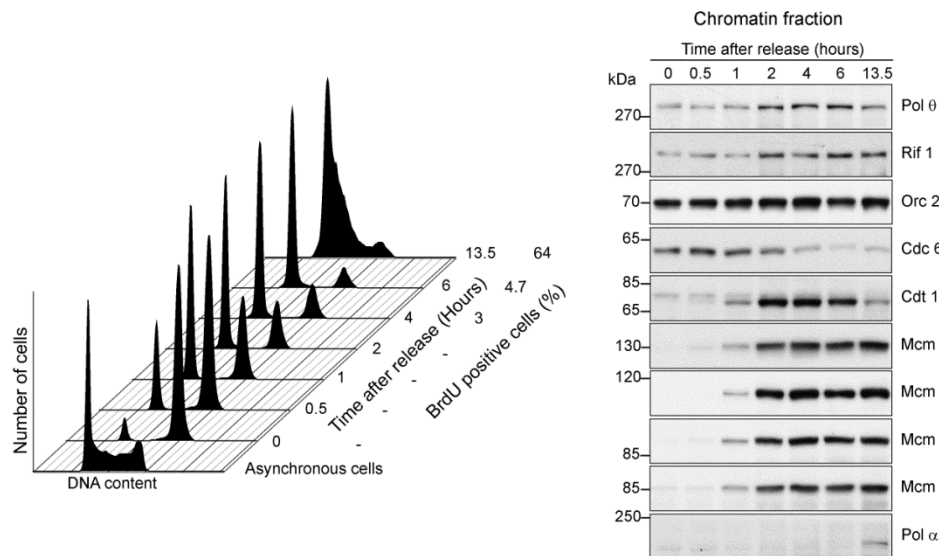
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Fernandez-Vidal, A. *et al.* A role for DNA polymerase θ in the timing of DNA replication. *Nat. Commun.* **4**:4285 doi: 10.1038/ncomms5285 (2014).

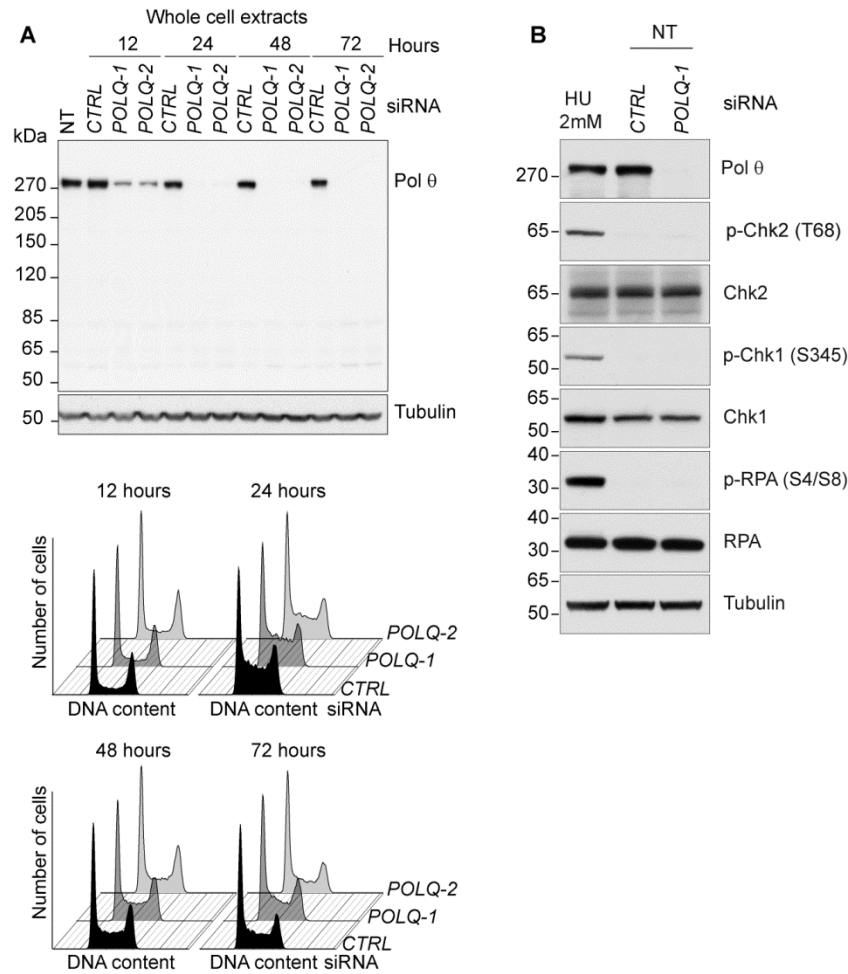
Supplementary Figures

Supplementary Figure 1



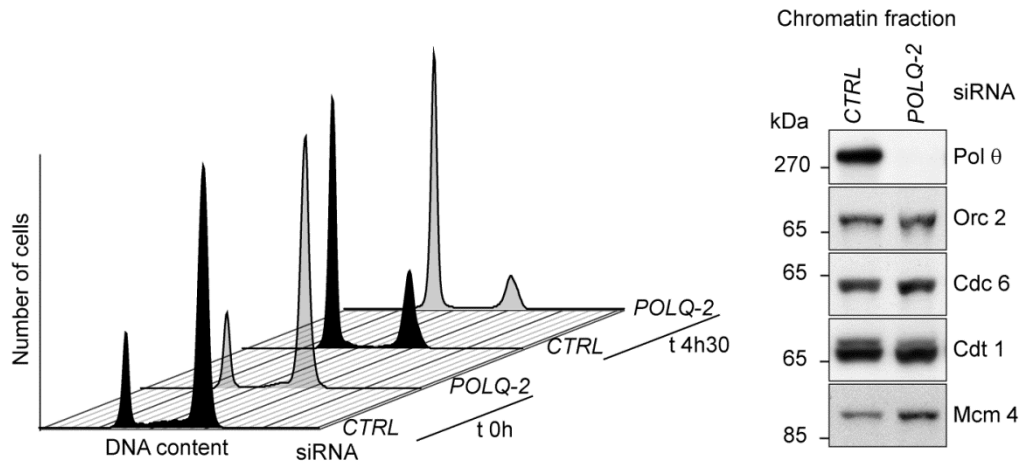
Supplementary Figure 1: Further demonstration that Pol θ is recruited onto chromatin in early G1 in human cells and controls the recruitment of Mcm onto chromatin. Human RKO cells were synchronized in mitosis using the Cdk1 inhibitor RO-3306 then released into fresh medium for the indicated time (0, 30min, 1h, 2h, 4h, 6h and 13h30). Cell cycle progression was followed by FACS analysis after BrdU incorporation and Propidium iodide staining (**left panel**). After subcellular fractionation, chromatin fractions were analyzed at the same time points throughout the cell cycle by western blotting with antibodies against Pol θ , Rif1, the pre-replication complex and the replicative DNA polymerase α (**right panel**). Representative of three independent experiments.

Supplementary Figure 2



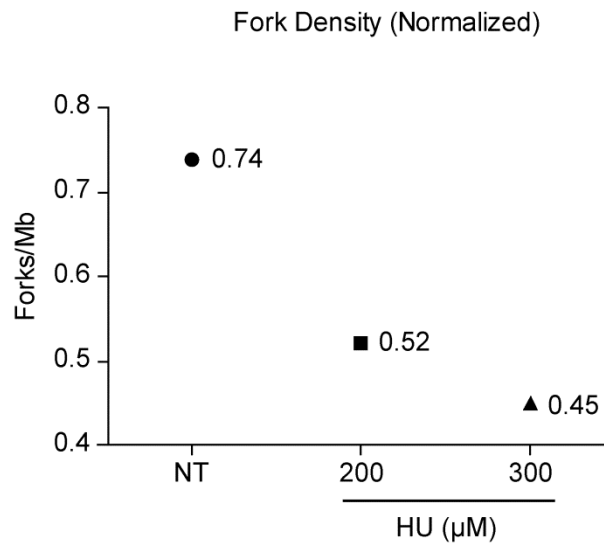
Supplementary Figure 2: Pol θ depletion has no impact on cell cycle progression and DNA damage signalling. A: Kinetics of Pol θ depletion and cell cycle analysis on asynchronous cells. 12, 24, 48 and 72h after transfection of RKO cells with control (*CTRL*) or *POLQ-1* or *-2* siRNA (*POLQ-1*, *POLQ-2*), western blotting was used to detect Pol θ in whole cell extracts (**upper panel**). The impact of Pol θ depletion on the cell cycle was also analyzed by propidium iodide staining coupled with FACS analysis (**lower panel**). **B: DNA damage signaling is not activated following Pol θ depletion.** 24h after transfection of RKO cells with control (*CTRL*) or *POLQ-1* siRNA, whole cell extracts were prepared and subjected for western blotting with the indicated antibodies. As a positive control, cells were treated with HU (2 mM for 3 hours).

Supplementary Figure 3



Supplementary Figure 3: Further demonstration that MCM loading onto chromatin is enhanced in G1 after Pol θ depletion. 26 hours after Control (*CTRL*) or *POLQ-2* siRNA transfection, human RKO cells were synchronized in mitosis using the Cdk1 inhibitor RO-3306 then released into fresh medium and collected in mitosis (t0h) or 4h30 later in G1. The cell cycle analysis of mock- (*CTRL* siRNA) and Pol θ- (*POLQ-2* siRNA) depleted cells was performed by propidium iodide staining and FACS analysis (**left panel**). 4h30 after the release, cell fractionation was performed and processed for immunoblotting with the indicated antibodies (**right panel**).

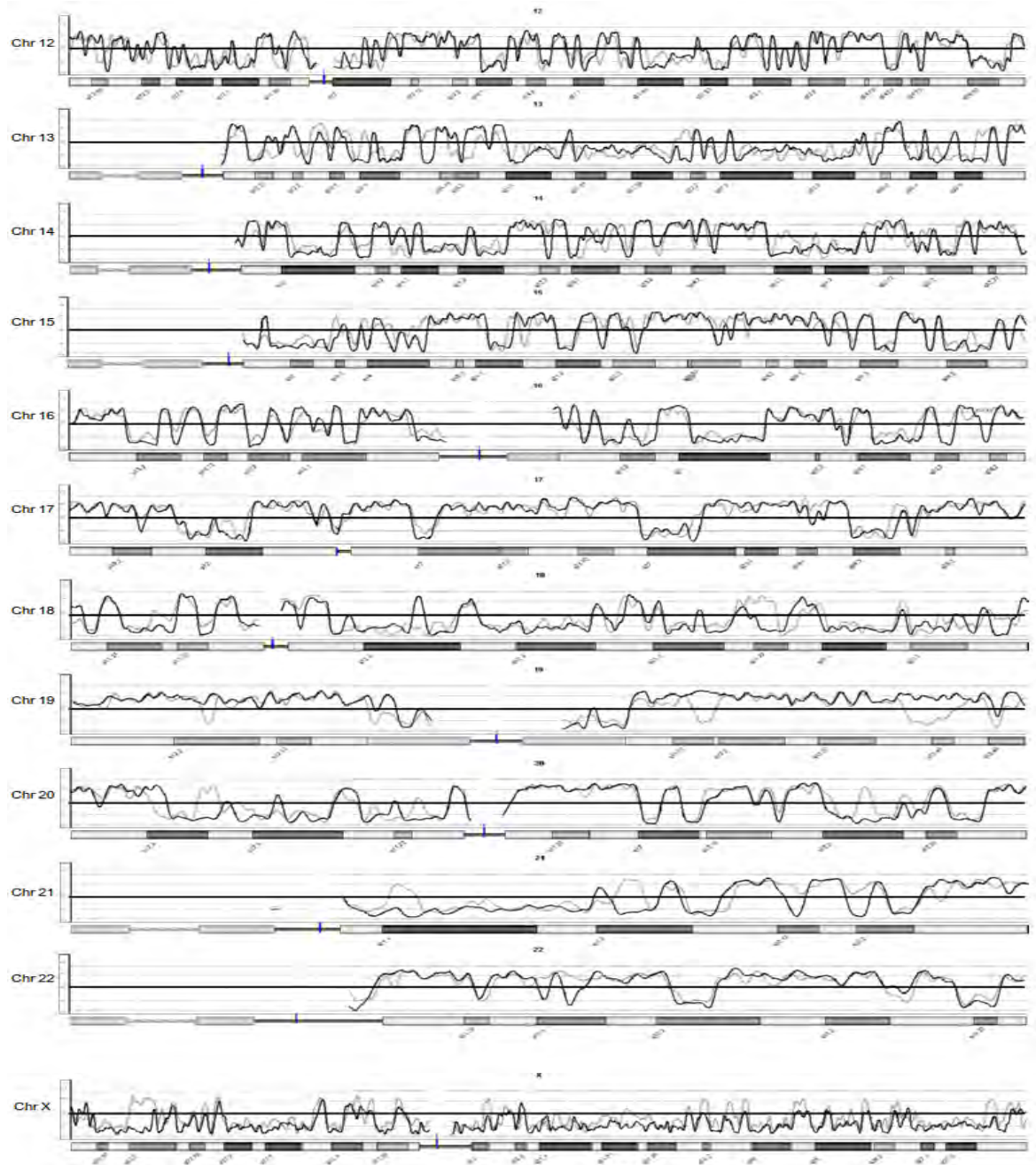
Supplementary Figure 4



Supplementary Figure 4: Positive control showing a modification of origin activity in RKO cells by DNA Combing after replication stress. RKO cells were treated or not with low doses of HU (200 and 300 μM for 20 hours). DNA replication was then monitored on single molecules spread by DNA combing after sequential pulse-labeling with IdU and CldU as described in Methods. The observed inhibition of fork density reveals inhibition of distant origins in HU-treated cells.

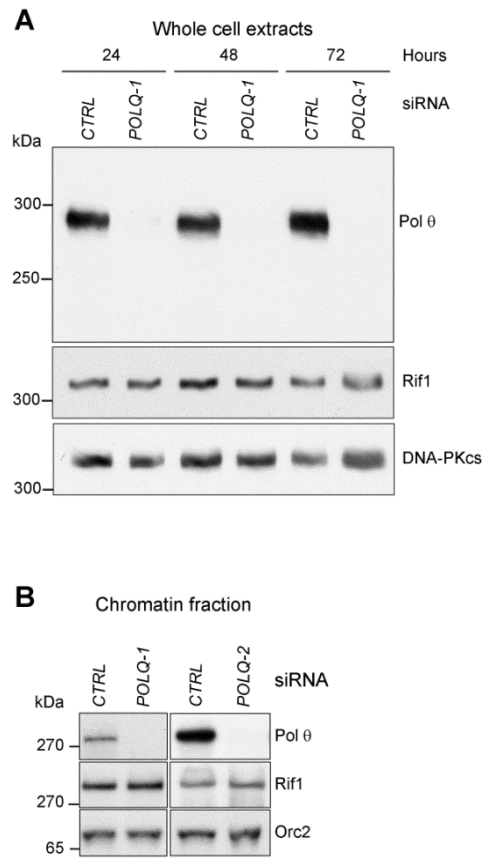
Supplementary Figure 5





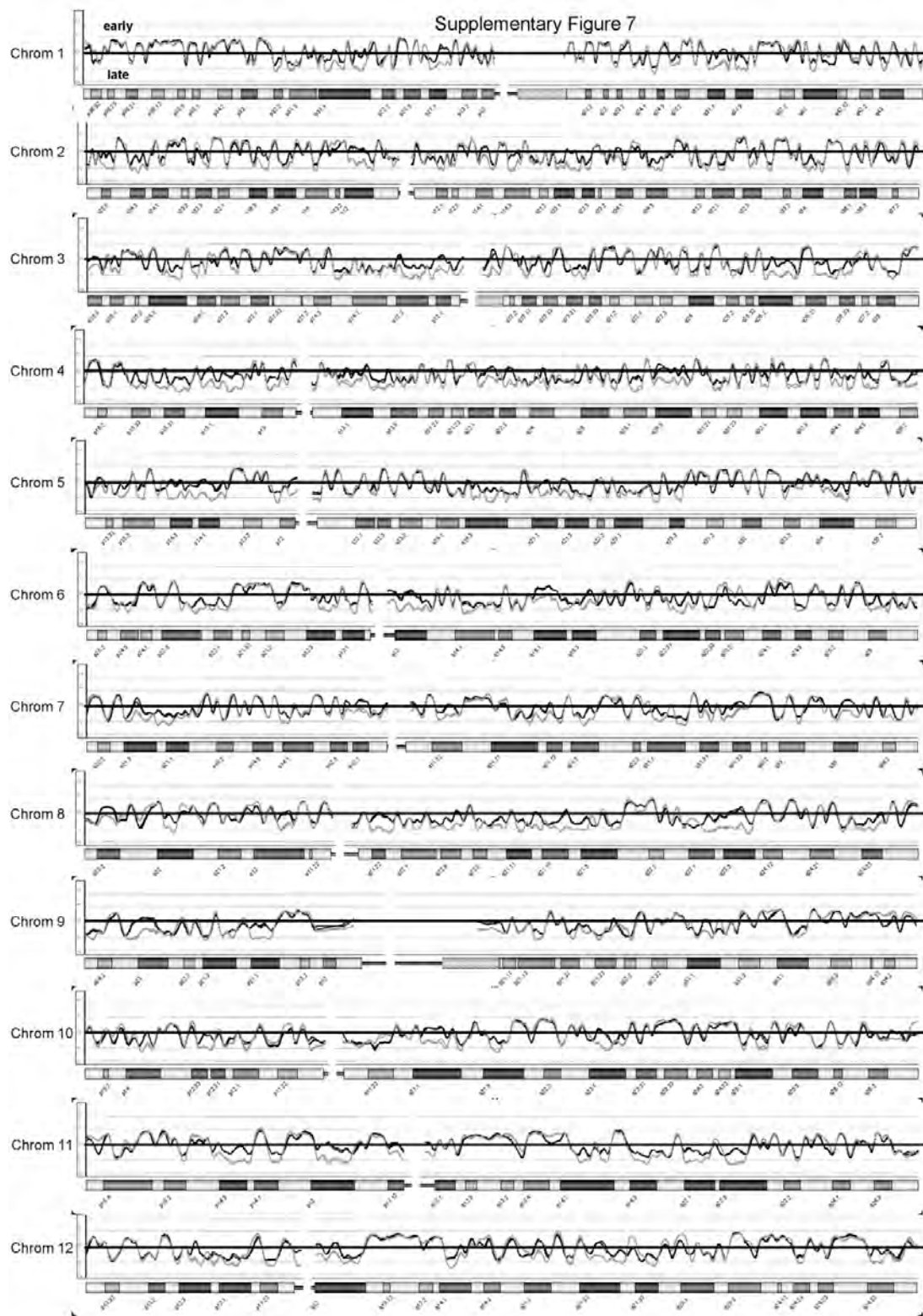
Supplementary Figure 5: Microarray profiles of the replication timing of the whole genome following Pol θ depletion. 24hours after transfection of RKO cells with either *POLQ-1* or *CTRL* siRNA, the replication timing was analysed. Black lines represent cells deficient in Pol θ and grey lines represent control cells. The horizontal line represents the zero value. Above this line replication was early during S-phase and below the line replication was late. All 23 chromosomes are shown.

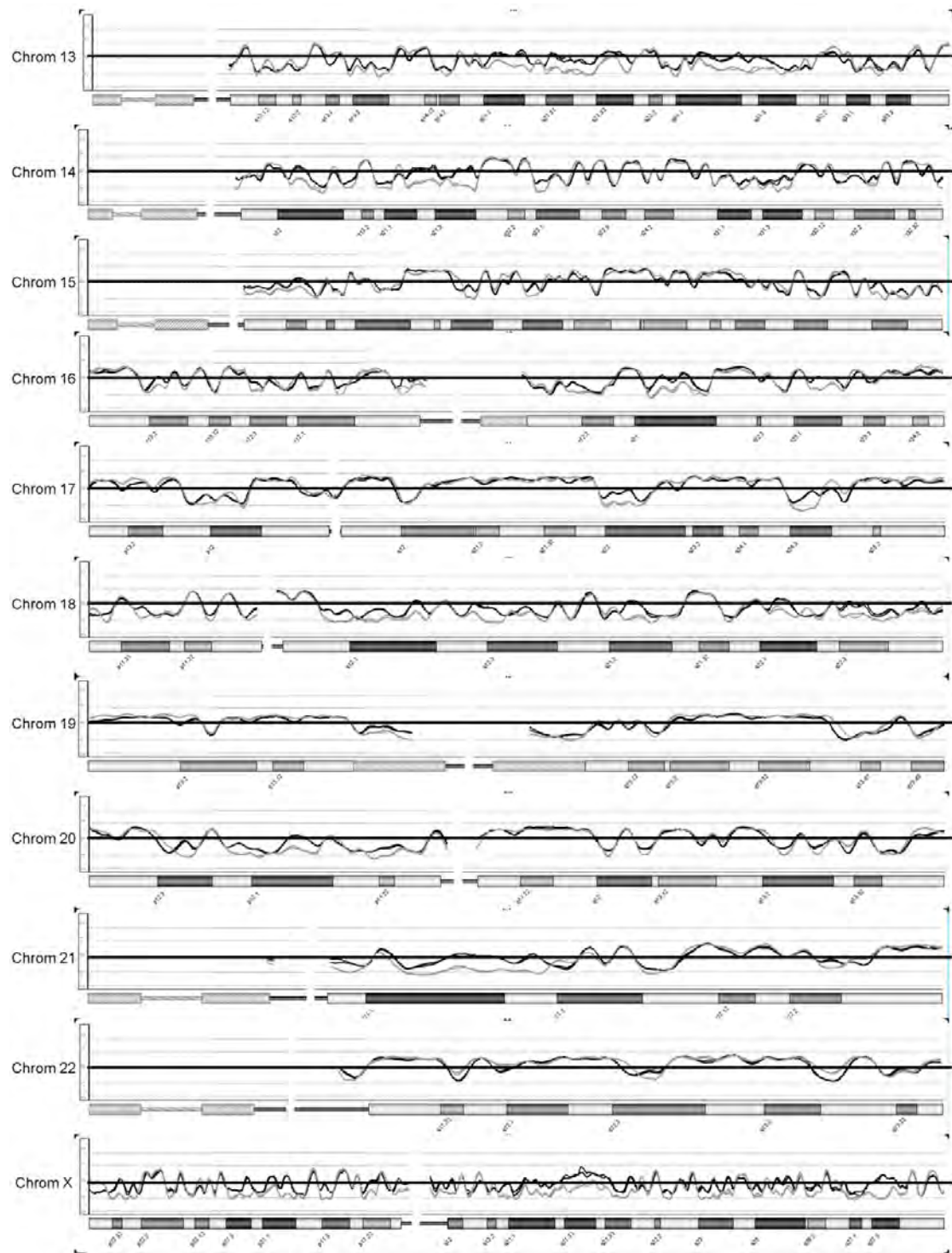
Supplementary Figure 6



Supplementary Figure 6: Pol θ does not affect the level of Rif1 or its recruitment to chromatin. **A:** 24, 48, and 72h after transfection with POLQ1 siRNA, the levels of Pol θ, Rif1, and DNA-PKcs (used as loading control) were analyzed by western blotting. **B:** 30 hours after transfection with control (CTRL) or POLQ (POLQ-1 or POLQ-2) siRNA, RKO cells were synchronized in G2/M with nocodazole (left part) or RO 3306 (right part), then released into fresh medium and collected 4h30 later in G1. Chromatin fractions were processed for immunoblotting with the Pol θ and Rif1 antibodies. Orc2 immunodetection was used as a chromatin fraction control and a loading control.

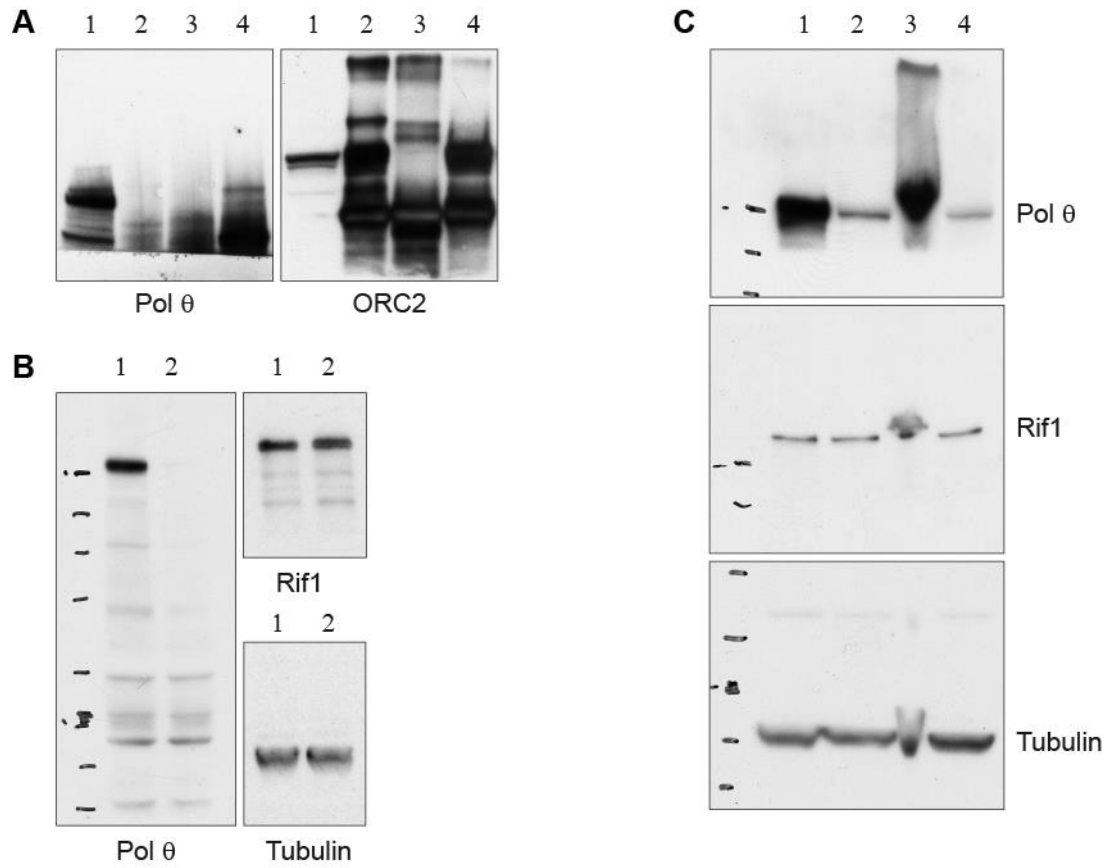
Supplementary Figure 7





Supplementary Figure 7: Microarray profiles of the replication timing of the whole genome in cells overexpressing Pol θ. Black lines represent cells overexpressing Pol θ and grey lines represent control (CTRL) cells. The horizontal line represents the zero value. Above this line replication is early during S-phase and below the line replication is late. All 23 chromosomes are shown.

Supplementary Figure 8



Supplementary Figure 8: Full blots relative to figures of the main manuscript. A: Full blot of Figure 2A: IP ORC2 (lane 1: Input, lane 3: IP IgG, lane 4, IP ORC2). **B:** Full blot of Figure 4A (lane 1: CTRL siRNA, lane 2: *POLQ-1* siRNA). **C:** Full blots of Figure 6A (lane 1: Pol θ overexpressing cells, lane 2: Control cells).

Supplementary Tables

Supplementary Table 1: Characteristics of the domains EtoL

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr6	126837441	126842805	5364	34	0	3.0947054437	0
chr13	39871564	39879039	7475	44	133.779264214	2.0602006689	0
chr12	42863878	42871642	7764	37	128.7995878413	3.5162287481	0
chr20	59703449	59711329	7880	56	126.9035532995	0	0,0117839014
chrX	44376495	44385080	8585	45	0	5.2533488643	0,0080243318
chr6	89601731	89610955	9224	35	108.4128360798	24.143538595	0
chr14	57378243	57388071	9828	39	0	11.63003663	0
chr18	72538855	72548820	9965	44	0	15.8354239839	0
chr11	102540720	102551227	10507	33	95.1746454744	26.3062720091	0
chr21	25680520	25692921	12401	38	80.6386581727	8.5396339005	0
chr4	185297072	185311144	14072	48	71.0631040364	12.464468448	0,0039972996
chr18	75338221	75353107	14886	61	67.1772134892	0	0,0210341767
chr11	82744349	82761437	17088	40	0	14.84082397	0
chr4	115914218	115931346	17128	35	0	59.0611863615	0
chr9	20467426	20484726	17300	36	57.8034682081	3.8728323699	0,0046563905
chr2	236504430	236521958	17528	47	57.0515746235	3.0465540849	0,0029132719
chr3	181002007	181020219	18212	40	54.9088513068	11.2563145179	0,0038436196
chr11	102577999	102597372	19373	34	51.6182315594	16.6726887937	0
chr7	111031308	111050894	19586	42	0	6.4689063617	0,0035253558
chr19	21719412	21740222	20810	41	48.0538202787	17.8183565593	0,0028129066
chr4	178749758	178771042	21284	37	0	9.4108250329	0,0026666396
chr4	174068603	174092010	23407	39	42.722262571	26.2784637074	0,0019717967
chr9	103127012	103151694	24682	40	40.5153553197	6.2312616482	0
chr2	67843620	67871208	27588	40	0	15.5864868784	0,0025373351
chr7	71580798	71608619	27821	41	0	28.460515438	0
chr15	71055370	71083370	28000	40	0	24.9642857143	0
chr13	39895488	39924340	28852	43	34.6596423125	0.6169416332	0,0023375108
chr7	102963127	102993969	30842	38	32.4233188509	4.5327799754	0
chr15	53349787	53380879	31092	43	32.1626141773	22.7100218706	0,0021691065
chr9	21189335	21220800	31465	37	127.1253774035	38.9162561576	0
chr9	103405035	103437915	32880	39	30.4136253041	8.3667883212	0,002027575
chr12	95835594	95869989	34395	34	29.073993313	17.0722488734	0
chr6	20139380	20177392	38012	44	0	4.214458592	0,0101642544
chr9	20510469	20548803	38334	39	26.0865028434	10.6980748161	0,0015384348

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr3	116133980	116172846	38866	35	25.7294293213	31.387330829	0
chr5	90996627	91035655	39028	36		14.2308086502	0
chr12	9373242	9412611	39369	45		17.7550864894	0,006152613
chr1	99713108	99753930	40822	39		11.2831316447	0,0015074828
chr10	11492484	11533423	40939	47		14.939299934	0,0051451745
chr2	206964870	207006243	41373	42		23.7763758973	0,0039708436
chr10	87927977	87969869	41892	49	23.8709061396	13.981189726	0,0113508595
chr11	106901464	106944787	43323	36	23.0824273481	13.2446968123	0
chr10	26896484	26940260	43776	40	45.6871345029	15.7118055556	0,0014848319
chr6	109969009	110013908	44899	37	22.2722109624	42.6312390031	0,0019262453
chr22	21704970	21750471	45501	50	21.9775389552	14.8941781499	0,0131865234
chr3	180307789	180353704	45915	41	21.7793749319	53.3028422084	0,0078618231
chr12	60723818	60770187	46369	37	21.5661325455	11.276930708	0
chr4	78703979	78751228	47249	41	21.1644690893	9.4795657051	0,0011356544
chr14	36561053	36609153	48100	38	20.79002079	7.1891891892	0
chr1	3034756	3086412	51656	57	19.3588353725	1.4461050023	0,0277137012
chr4	111318442	111370777	52335	43	19.1076717302	1.8056749785	0,0051990642
chr12	14429962	14482466	52504	35	19.046167911	11.5781654731	0
chr4	186342845	186395887	53042	44	37.7059688549	4.8207081181	0,0041562261
chr21	32354552	32407616	53064	44	18.8451680989	14.0641489522	0,0010279183
chr8	33662401	33715832	53431	44		15.6968800883	0,0044237172
chr12	29249554	29303876	54322	39	18.408747837	2.9214682817	0
chr11	106968853	107023439	54586	39	18.319715678	9.4035100575	0,0010803935
chr1	227278990	227334586	55596	46		11.1860565508	0,0046922362
chr13	105942094	105998592	56498	41	35.3994831675	1.3150907997	0,0056552833
chrX	3628479	3686168	57689	46	17.3343271681	12.3940439252	0,0036929654
chr3	115330457	115388705	58248	42	17.1679714325	10.4724625738	0,0017167971
chr3	105911706	105970999	59293	34		10.9962390164	0,0011904986
chr2	183996857	184059048	62191	38		32.2120564069	0
chr1	58475086	58537413	62327	41	16.0444109295	18.4590947743	0,0020349009
chr13	105620049	105682403	62354	38		23.042627578	0
chr12	76077747	76140719	62972	38		24.7046306295	0
chr4	6276525	6341969	65444	53	15.2802395942	3.5205672025	0,0192012067
chr7	97260406	97326166	65760	39	15.2068126521	29.151459854	0,0012087466
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr1	3978378	4044602	66224	49		21.3215752597	0,0175347984
chr3	159905167	159972657	67490	44	14.8170099274	8.5479330271	0,0050849284
chr5	63895714	63967696	71982	39	13.8923619794	14.1452029674	0,0014604791
chr10	97733519	97806853	73334	39	27.2724793411	17.1952982246	0,0018181653
chr1	86633414	86707165	73751	40	40.6774145435	17.0289216417	0,0039999458
chr2	65862537	65936461	73924	40		8.3044748661	0,0029760294
chr12	29330185	29405940	75755	36	26.4008976305	15.7270147185	0,0023467465
chr4	100286292	100362808	76516	39	26.1383240107	31.8325578964	0
chr7	89505851	89582396	76545	37		33.8859494415	0
chr1	190689490	190766468	76978	37		30.3658188054	0,0018608335
chr1	34883375	34964118	80743	45		13.2816467062	0,0036329259
chr4	78782864	78866330	83466	38	11.9809263652	23.5748688089	0,0016079664
chr13	94620850	94705278	84428	43	11.8444118065	10.2004074478	0,0046826744
chr22	43752832	43837288	84456	52	11.8404849863	4.3904518329	0,0104287349
chr2	135168721	135256264	87543	43	11.4229578607	30.4319020367	0,0007438205
chr6	138636960	138724907	87947	42	11.3704844963	14.639498789	0,0016785001
chr1	34010936	34102632	91696	46	10.9056011167	14.1434740883	0,0029397707
chr18	47107705	47202562	94857	40		17.6296952254	0,0020030151
chr3	73330966	73426982	96016	42		13.8331111481	0,0016118345
chr8	92115920	92213566	97646	37	30.7232247097	9.9614935584	0,0005812502
chr10	31508437	31607706	99269	40		32.4481963151	0,0029717233
chr21	40009793	40109324	99531	43	10.0471209975	16.9042810783	0,0032010595
chr13	23916695	24017650	100955	43	9.9054033976	17.382001882	0,0051139525
chr6	109842223	109943482	101259	45	39.5026614918	4.0065574418	0,0099195572
chr11	60172766	60274080	101314	42	19.7406084056	23.325502892	0,0009870304
chr2	114390907	114495570	104663	40	9.5544748383	20.0653526079	0,0017436917
chr13	24149012	24254079	105067	46	19.0354726032	6.4282790981	0,010159149
chr7	155994634	156100148	105514	44	9.4774153193	17.6696931213	0,0050618014
chr5	106216777	106324444	107667	36		19.4600016718	0,001264186
chr14	69344719	69453529	108810	44	9.190331771	13.0328094844	0,0043862947
chr10	90367654	90478944	111290	37	17.9710665828	25.8396980861	0
chr10	42802131	42913430	111299	52	8.9848066919	22.442250155	0,0206304984
chr7	152195577	152307923	112346	46		19.8164598651	0,0117842473
chr10	16633961	16747751	113790	41	8.7881184638	16.8564900255	0,001328935
chr21	26207581	26321945	114364	40	8.7440103529	11.6111713476	0,0011804414

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr4	95255770	95374160	118390	39	8.4466593462	19.9535433736	0,0015377252
chr9	95105915	95225103	119188	44	8.3901063865	29.9921133	0,0032797689
chr1	243534366	243655241	120875	45	8.2730093071	5.7141675284	0,00292313
chr13	52058970	52185060	126090	40	31.7233721945	11.5544452375	0,0017447855
chr15	99968821	100094929	126108	42	15.8594220827	25.1427347987	0,0030208423
chr12	14505321	14632852	127531	40	15.6824615192	12.1876249696	0,0019014985
chr4	186147054	186275952	128898	43	7.7580722742	7.8922869245	0,0033377753
chr15	71160626	71289891	129265	37	7.7360461068	25.8631493444	0,0009199622
chr10	123360147	123492793	132646	46	7.5388628379	8.9961250245	0,0053591481
chr5	107299574	107432931	133357	36	7.4986689863	14.5526669016	0
chr3	180147708	180281871	134163	39	7.4536198505	27.3875807786	0,0009555923
chr9	91067400	91204052	136652	45	21.9535755057	14.1212715511	0,0050574533
chr7	89912037	90055271	143234	39	0	22.5951938785	0,0018975583
chr4	104323368	104470296	146928	38	6.8060546662	41.3426984646	0
chr2	118452204	118599560	147356	41	13.5725725454	5.0137082983	0,0036745257
chr20	13700083	13850277	150194	41	19.9741667443	8.691425756	0,0020948516
chr9	20240034	20397112	157078	40	6.3662638944	5.9575497524	0,002339602
chr12	70693541	70852772	159231	39	6.280184135	15.8995421746	0,0004830911
chr5	94008043	94172523	164210	37	12.1795262164	10.2990073686	0,0014812937
chr19	61428487	61594937	166450	43	18.0234304596	19.88164614	0,0024310674
chr3	106009505	106185010	175505	35	0	22.1748668129	0,000472107
chr4	80774408	80952255	177847	35	0	37.7571733006	0,0008193238
chr1	117348163	117526466	178303	42	22.4337223715	7.3139543362	0,0013086338
chr12	19262385	19444265	181880	40	5.4981306356	5.6713217506	0,0016219485
chr2	236013051	236197149	184098	48	5.4318895371	7.859441602	0,0066087989
chr18	72290844	72476474	185630	45	10.7741205624	13.7084522976	0,007912993
chr12	70015849	70204542	188693	38	5.2996136582	20.0913653395	0,0016596159
chr12	42212198	42402444	190246	37	5.2563523018	31.0009146053	0,0015200803
chr12	19538137	19730047	191910	40	5.2107758845	21.2229691001	0,001732583
chr6	87697052	87903306	206254	39	9.6967816382	14.815712665	0,0028095803
chr8	68287813	68496980	209167	39	9.5617377502	15.8772655342	0,001017467
chr4	81628004	81843571	215567	35	4.6389289641	27.3775670673	0,0003578602
chr14	71065782	71285674	219892	40	4.5476870464	15.3757299038	0,0023989049
chr8	1989140	2209311	220171	46	4.5419242316	7.5118884867	0,0040186156
chr9	19523171	19746191	223020	40	4.483902789	17.7674648014	0,0016254148
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr4	90220484	90443603	223119	39	8.9638264782	15.5419305393	0,0011836848
chr7	7301602	7528137	226535	39	4.4143289116	24.6045864877	0,0008262718
chr7	131923825	132156153	232328	44	8.6085189904	11.8255225371	0,0062216115
chr15	47423142	47655589	232447	36	4.3020559525	36.9809031736	0,0005497071
chr13	101859950	102097643	237693	38	4.2071074874	25.7634848313	0,0012842749
chr9	76662741	76917086	254345	40	19.6583380841	22.0421081602	0,002673534
chr5	177614460	177871940	257480	50	3.8837967998	14.7514370048	0,0098337735
chr2	118157131	118422181	265050	41	7.5457460856	29.765704584	0,0010122342
chr4	103221416	103491466	270050	39	3.7030179596	18.6772819848	0,000227878
chr7	89601591	89873674	272083	38	18.3767453314	21.1211284792	0,0011509646
chr3	19878220	20153553	275333	41	10.8958969684	11.3088514635	0,0013819186
chr15	47927378	48207020	279642	38	3.5760007438	21.9169509587	0,0027949269
chr4	115960412	116244465	284053	35	3.5204697715	16.1008684999	0,0008952052
chr9	20883453	21168081	284628	37	17.5667889315	27.1385808845	0,0008166183
chr4	126364222	126653498	289276	35	3.4569062072	9.7920325226	0,0004543362
chr12	64906813	65197459	290646	40	10.3218348094	15.9021627685	0,0012730263
chr6	101425066	101717320	292254	37	3.4216811404	31.4722809611	0
chr12	10736448	11030716	294268	37	20.3895768483	28.5668166433	0,0016440244
chr19	20001470	20299918	298448	42	13.4026698118	12.7147777837	0,0011966669
chr21	32015575	32317196	301621	45	6.6308380385	11.4570935048	0,0050762749
chr19	55918797	56220927	302130	50	56.2671697614	6.1122695528	0,0157349485
chr5	179794252	180097347	303095	47	9.8978868012	19.5681222059	0,0117581279
chr4	113727172	114035284	308112	41	6.4911460768	7.4171080646	0,000506626
chr18	57584631	57893021	308390	41	6.4852945945	13.3434936282	0,0046504308
chr1	86728636	87039038	310402	38	12.8865149065	28.9060637496	0,0010173564
chr3	159495038	159805828	310790	36	6.4352134882	22.2896489591	0,0021182578
chr6	24380303	24700623	320320	41	15.6093906094	11.5206668332	0,0016370824
chr9	87332774	87653240	320466	38	6.2409116724	21.2190996861	0,0016916155
chr4	78932007	79252515	320508	39	9.3601407765	19.7985697705	0,001864028
chr7	147702365	148025704	323339	43	6.1854586054	17.8413368013	0,0003452349
chr8	86207163	86534713	327550	39	12.2118760495	18.3434590139	0,0008297813
chr6	97596159	97923892	327733	37	6.1025285827	11.8495848755	0,0019791985
chr6	89641265	89972710	331445	41	15.0854591259	14.153177752	0,0021340406
chr5	94200402	94538826	338424	37	2.9548731768	17.2608916625	0,0009822957
chr3	178177064	178523142	346078	41	2.8895220153	11.6930287392	0,0019874273

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr3	87953230	88299787	346557	37	11.5421128415	24.8643080359	0,001076224
chr7	55768247	56115221	346974	45	20.174422291	15.0057353001	0,0033944266
chr1	87123525	87471811	348286	39	8.6136106533	8.6184916994	0,0030110827
chr18	49934991	50289530	354539	38	11.282256677	18.4312586203	0,0018333667
chr4	24385412	24741196	355784	44	11.2427765161	9.9338362602	0,004567378
chr12	42457755	42814908	357153	38	8.3997614468	16.2174194253	0,0014589059
chr14	56661989	57026163	364174	38	8.2378203826	17.1110513106	0,0021317167
chr10	97824755	98189960	365205	42	13.6909407045	16.1599649512	0,0035922421
chr9	85225118	85611155	386037	41	10.3617010805	18.0933433842	0,0014784378
chr10	22497393	22884456	387063	40	10.3342349953	16.5629367829	0,0044889333
chr19	42641137	43028982	387845	41	20.6267967874	18.4921811548	0,0014401026
chr3	109526528	109918060	391532	37	7.6622089638	19.6875351185	0,0009733076
chr1	243106563	243507902	401339	44	4.9833183419	13.7357695116	0,0033467513
chr3	122558968	122963787	404819	40	14.8214387171	23.0885902095	0,0013154027
chr5	41713438	42123863	410425	37	7.3094962539	27.7373454346	0,0014289736
chr7	26379776	26791320	411544	41	12.1493692048	9.6538887701	0,0033129256
chr8	81018410	81432474	414064	41	4.8301711813	13.1593183662	0,0021146725
chr1	61081276	61496565	415289	40	2.4079616845	5.3483236975	0,004653386
chr7	63523852	63939743	415891	41	12.0223808642	21.4779834139	0,0017828311
chr3	156603423	157022673	419250	40	4.7704233751	21.4015503876	0,0012820513
chr3	98981714	99406033	424319	37	16.4970222875	17.3169242952	0,0010318601
chr7	25930676	26362247	431571	43	11.5855791979	10.7219901245	0,0035942239
chr2	156726566	157167846	441280	38	4.5322697607	16.0363034808	0,0041327144
chr3	25550094	25997193	447099	39	8.9465644074	15.8542067864	0,0018007828
chr4	35882	483423	447541	41	11.1721607629	27.7520942215	0,0013352095
chr3	103360326	103815470	455144	37	2.1971068497	19.5509553021	0,0009085334
chr13	45416528	45878360	461832	39	8.6611581701	19.2754075075	0,0009993644
chr7	81337417	81802509	465092	35	2.1501122359	10.7862530424	0,0009706221
chr3	162205828	162671543	465715	38	8.588943882	23.7683991282	0,0006893758
chr14	36741208	37207185	465977	37	4.2920573333	30.0480495818	0,0019720263
chr10	118679339	119158932	479593	42	10.4255066275	17.7661058439	0,0051035337
chr2	151840463	152325930	485467	39	8.2394889869	11.2909837332	0,0018274764
chr16	51771379	52262967	491588	40	6.1026713427	17.547214334	0,001835887
chr8	9200933	9709111	508178	40	1.9678144272	12.2972659186	0,0018251479
chr7	86978375	87488740	510365	37	5.8781460327	29.6050865557	0,0009955779
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr10	30617288	31133027	515739	44	7.7558610072	10.3030796585	0,0047019907
chr7	149453224	149986532	533308	43	16.8758016006	27.7631312487	0,0069334689
chr1	63211717	63746851	535134	40	5.6060724977	12.6435621732	0,0034710932
chr19	9089788	9634799	545011	43	23.8527295779	15.8156440879	0,0013483833
chr6	26245517	26797314	551797	43	52.5555593814	8.0013120767	0,0036034485
chr13	26919442	27485233	565791	43	15.9069338325	15.1336801045	0,0063833251
chr14	59453522	60020302	566780	37	7.057412047	30.1533222767	0,0009203247
chr12	22369323	22937603	568280	37	5.2790877736	19.9707890477	0,0016122619
chr5	79879949	80452676	572727	40	5.2380977324	19.2253901073	0,0023789694
chr5	1131749	1712250	580501	55	15.5038492612	9.528838021	0,0218275405
chr13	24406647	24990377	583730	42	10.2787247529	15.867267401	0,0027246778
chr20	5532318	6140222	607904	43	13.1599726273	10.8977733326	0,0021537978
chr5	96788425	97405778	617353	37	0	29.912060037	0,0001357145
chr4	106308209	106955663	647454	37	9.2670676218	24.5340981753	0,0010477631
chr12	40710016	41362758	652742	40	7.659994301	12.6509401877	0,0016660488
chr6	135149878	135829104	679226	39	5.8890560726	11.2854926048	0,0022763467
chr12	4112304	4800329	688025	42	15.9877911413	19.9498564732	0,0057687549
chr4	16929123	17624310	695187	41	10.0692331704	14.939865101	0,0019050849
chr5	78788466	79498375	709909	41	8.4517874826	12.4868116899	0,0029478186
chr9	6202886	6913587	710701	42	7.0353073937	16.7732984757	0,0034607012
chr4	67788683	68499487	710804	37	9.8480031063	23.5085902724	0,0003650225
chr19	41872197	42610151	737954	43	17.6162741851	22.6489727002	0,0012920702
chr6	136722847	137487752	764905	40	7.8441113602	13.1212372778	0,0020035168
chr12	73170183	73938013	767830	35	2.6047432374	20.7650130888	0,0005656014
chr18	27142010	28101721	959711	40	12.503764154	11.0766678719	0,0018260706
chr4	125223012	126259284	1036272	36	0.9649976068	21.1905754474	0,0004556933
chr1	61529637	62580402	1050765	42	4.7584379	8.9421516705	0,0020778512
chr6	26853170	27907672	1054502	41	18.0179838445	16.6514620171	0,0017439743
chr13	47234365	48289520	1055155	40	6.6340964124	22.4763186451	0,0015637513
chr6	6832950	8068275	1235325	44	8.0950357193	10.31841823	0,00428117
chr6	27968908	29292243	1323335	40	21.1586635281	20.1286144476	0,0018910556
chr19	56242634	58624819	2382185	45	29.3847874955	10.5888921305	0,0072426505

Supplementary Table 2: Characteristics of the domains LtoE

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr9	117204475	117212236	7761		38	128.84937506	6.8032470043
chr2	228288125	228297942	9817		45	101.86411327	0
chr3	55458392	55468330	9938		42	0	13.8257194607
chr6	71844384	71854400	10016		39	0	4.0135782748
chr2	215980142	215992397	12255		38	81.599347205	4.2676458588
chr5	155135168	155150066	14898		37	0	8.860249698
chr4	141836693	141852171	15478		42	64.607830465	9.8333117974
chrX	41438705	41454396	15691		38	63.730801096	9.8655280097
chr21	16082669	16099445	16776		33	59.608965186	19.6292322365
chr12	102433842	102451090	17248		41	0	3.9946660482
chr7	126730581	126748425	17844		37	0	52.477023089
chr7	135353693	135372541	18848		40	0	2.5042444822
chr1	77383373	77402264	18891		36	52.935260176	28.9767614208
chr1	95626247	95646480	20233		38	0	27.3810112193
chr12	97047158	97067610	20452		39	0	0.488949736
chr6	18675106	18695657	20551		39	48.659432631	21.2301104569
chr7	157324610	157346118	21508		54	46.494327692	1.1391110285
chr1	170335776	170357888	22112		38	45.224312590	7.4574891462
chr11	28306229	28328613	22384		35	44.674767691	21.8504288778
chr5	168941650	168964385	22735		37	43.985045084	3.633164724
chr15	91627408	91651439	24031		43	0	17.2485539511
chr10	80346469	80370638	24169		50	0	5.0684761471
chr5	38067580	38091928	24348		41	0	5.7704944965
chr10	99303217	99330057	26840		48	74.515648286	3.0365126677
chr22	26295996	26324575	28579		49	0	2.4983379405
chr1	18079338	18108535	29197		48	0	15.0803164709
chr21	29994506	30023991	29485		37	33.915550275	7.7700525691
chr14	62714174	62746183	32009		40	31.241213406	5.9045893343
chr14	52891007	52924255	33248		40	0	12.5090230991
chr8	123433977	123467320	33343		37	0	22.9343490388
chr2	148154471	148187869	33398		40	0	5.9823941553
chr14	95525976	95560117	34141		41	0	41.9671362877
chr11	110434442	110472521	38079		39	0	14.3097245201
chr21	16334316	16373603	39287		35	25.453712424	15.2111385446
chr5	35237418	35276761	39343		40	25.417482144	11.0972727042
chr10	80395659	80435399	39740		52	25.163563160	5.6919979869
chr8	32052275	32093185	40910		39	24.443901246	6.2723050599
chr7	131201126	131242744	41618		45	0	5.6850401269
chr16	654002	696450	42448		65	141.349415740	0.6572747833
chr5	75910722	75953438	42716		39	23.410431688	5.8221743609
chr17	28249104	28292825	43721		46	22.87230392	0.6953180394
chr18	18458071	18502577	44506		37	0	49.9213589179
chrX	14909253	14955154	45901		35	0	36.0253589246
chr18	66505317	66551418	46101		39	0	18.845578187
chr2	29353772	29400588	46816		46	21.360218728	3.5372522215
chr13	96896039	96944051	48012		37	20.828126307	20.3553278347
chr5	169483110	169533869	50759		45	0	11.2118048031
chr9	22545507	22596381	50874		36	0	43.4445885914
chr7	77681943	77735361	53418		39	18.720281552	21.0228761841
chr11	96028955	96084576	55621		36	0	22.5634202909
chr12	130241356	130299366	58010		52	17.238407177	7.8710567144
chrX	149978759	150036780	58021		43	0	32.810534117
chr22	33494228	33552797	58569		42	0	31.6703375506
chr5	10924380	10983173	58793		41	0	16.8098242988
chr7	3191827	3251321	59494		45	0	20.3415470468
chr3	55291127	55350892	59765		42	0	8.5116707103
chr15	55754443	55814304	59861		40	16.705367439	0.0593207598
chr21	38058184	38119675	61491		42	16.262542485	18.4433494333
chr9	75045167	75106817	61650		36	0	12.1054339011
chr4	157562613	157625602	62989		36	0	16.088523393

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC	
chr10	21383767	21447026	63259		40	15.808027316	24.7601131855	0,001027522
chr6	52005762	52070462	64700		40	15.455950541	14.4544049459	0,003593509
chr14	100917549	100982529	64980		42	0	33.6380424746	0,002564892
chr14	24314763	24381667	66904		40	14.946789429	5.8636254932	0,003811431
chr2	236917363	236987161	69798		45	14.327058081	26.5494713316	0,011907377
chr11	16563064	16633091	70027		39	0	37.0742713525	0,005931778
chr11	95031859	95103009	71150		38	0	17.650035137	0,001664386
chr16	2809231	2881285	72054		52	69.392400144	6.6339134538	0,018469054
chr17	14910874	14984137	73263		42	0	4.8237172925	0,004224831
chrX	133078001	133153090	75089		40	79.905179187	15.048808747	0,005193837
chr14	21748129	21825642	77513		40	0	19.2187117	0,000709558
chr1	90493740	90571267	77527		39	0	18.412940008	0,002116715
chr12	130430647	130509117	78470		53	0	8.6835733401	0,011445307
chr4	149340301	149419423	79122		38	12.638709846	5.4915194257	0,003758353
chr7	105830909	105912322	81413		39	0	29.1611904733	0,00075588
chr12	97888212	97971107	82895		39	12.063453766	18.9203208879	0,000556775
chr3	30518045	30601001	82956		38	0	16.4798206278	0
chr8	123524289	123607565	83276		40	0	18.7496997935	0,001651136
chr10	100924276	101009416	85140		39	11.745360582	40.7329105004	0,002800817
chr6	140332617	140417956	85339		37	0	12.5534632466	0,003103679
chr8	130525173	130611601	86428		39	0	19.2576479844	0,002195395
chr9	71483663	71571718	88055		36	11.356538527	25.8895008801	0,001545751
chrX	67093293	67181876	88583		38	11.288847747	57.6058611698	0,001930987
chr7	120216180	120307136	90956		37	10.994326929	9.3649676767	0,000653717
chr14	98659223	98751915	92692		50	10.788417554	5.4557027575	0,014801709
chr11	111831632	111926336	94704		42	0	15.8990116574	0,005480736
chr21	38985411	39080759	95348		43	10.487896966	6.3766413559	0,002219532
chr2	29422941	29518565	95624		44	10.457625706	6.037187317	0,001259669
chr15	27690169	27788551	98382		46	10.164460978	17.2429916042	0,005634647
chr12	5687067	5786793	99726		44	10.027475281	12.9805667529	0,003646355
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC	
chr2	224088708	224190340	101632		39	9.8394206541	12.033611461	0,002169718
chr21	30051585	30156956	105371		38	9.490277211	4.393049321	0
chr4	141561293	141668144	106851		38	18.717653554	31.0132801752	0,001329939
chr12	130026622	130133704	107082		52	9.3386376795	13.3215666499	0,011888804
chr6	154708444	154816351	107907		42	18.534478764	8.652821411	0,002471264
chr5	33621241	33730222	108981		40	9.1759113974	12.0672410787	0,00100935
chr4	169047365	169156523	109158		37	0	36.9226259184	0,000668508
chr3	36664591	36778031	113440		41	8.8152327221	20.0035260931	0,001634043
chr19	12498238	12615984	117746		46	25.478572520	9.3599782583	0,002141677
chr14	53520207	53638538	118331		38	0	38.4362508557	0,000444783
chr7	116867726	116986789	119063		36	8.3989148602	21.749829922	0,00072324
chr21	15777833	15899387	121554		38	0	17.4663112691	0,000476288
chr1	84037542	84159398	121856		37	8.206407563	13.2361147584	0,002861153
chr12	78060891	78183903	123012		36	8.1292881995	11.4541670731	0
chr5	33746151	33870779	124628		40	8.0238790641	16.4914786404	0,001444298
chr17	67905719	68033515	127796		47	0	5.4688722652	0,009140232
chr16	82106386	82234550	128164		43	7.802503043	13.651259324	0,000362907
chr9	108764605	108908837	144232		41	6.9332741694	12.3544012424	0,002333639
chr21	16500877	16645262	144385		34	6.9259272084	6.1661529937	0,001140741
chr11	48297770	48443761	145991		39	13.699474625	47.813221363	0,000456649
chr21	16174123	16320221	146098		37	6.8447206664	28.612301332	0
chr1	230300037	230450586	150549		42	0	13.8579465822	0,002008522
chrX	107103875	107256716	152841		40	26.170988151	33.5145674263	0,004171001
chr7	115230262	115384022	153760		35	6.5036420395	16.4672216441	0,001598038
chr2	1675048	1835556	160508		51	12.460438105	7.7616068981	0,010774614
chr5	135679332	135844388	165056		40	6.0585498254	23.6846888329	0,001272296
chrX	109764698	109933737	169039		39	5.9157945795	26.8979347961	0,003079247
chr6	46974023	47151440	177417		40	11.272876894	9.1327212161	0,002522306
chr3	135917277	136097185	179908		44	5.5583965138	17.0092491718	0,00789545
chr6	13138755	13319020	180265		41	5.5473885668	12.1532188722	0,004952059

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr2	224088708	224190340	101632	39	9.839420654	12.033611461	0,002169718
chr21	30051585	30156956	105371	38	9.490277211	4.393049321	0
chr4	141561293	141668144	106851	38	18.71765355	31.0132801752	0,001329939
chr12	130026622	130133704	107082	52	9.338637679	13.3215666499	0,011888804
chr6	154708444	154816351	107907	42	18.53447876	8.652821411	0,002471264
chr5	33621241	33730222	108981	40	9.175911397	12.0672410787	0,00100935
chr4	169047365	169156523	109158	37	0	36.9226259184	0,000668508
chr3	36664591	36778031	113440	41	8.815232722	20.0035260931	0,001634043
chr19	12498238	12615984	117746	46	25.47857252	9.3599782583	0,002141677
chr14	53520207	53638538	118331	38	0	38.4362508557	0,000444783
chr7	116867726	116986789	119063	36	8.398914860	21.749829922	0,00072324
chr21	15777833	15899387	121554	38	0	17.4663112691	0,000476288
chr1	84037542	84159398	121856	37	8.206407563	13.2361147584	0,002861153
chr12	78060891	78183903	123012	36	8.129288199	11.4541670731	0
chr5	33746151	33870779	124628	40	8.023879064	16.4914786404	0,001444298
chr17	67905719	68033515	127796	47	0	5.4688722652	0,009140232
chr16	82106386	82234550	128164	43	7.802503043	13.651259324	0,000362907
chr9	108764605	108908837	144232	41	6.933274169	12.3544012424	0,002333639
chr21	16500877	16645262	144385	34	6.925927208	6.1661529937	0,001140741
chr11	48297770	48443761	145991	39	13.69947462	47.813221363	0,000456649
chr21	16174123	16320221	146098	37	6.844720666	28.612301332	0
chr1	230300037	230450586	150549	42	0	13.8579465822	0,002008522
chrX	107103875	107256716	152841	40	26.17098815	33.5145674263	0,004171001
chr7	115230262	115384022	153760	35	6.503642039	16.4672216441	0,001598038
chr2	1675048	1835556	160508	51	12.46043810	7.7616068981	0,010774614
chr5	135679332	135844388	165056	40	6.058549825	23.6846888329	0,001272296
chrX	109764698	109933737	169039	39	5.915794579	26.8979347961	0,003079247
chr6	46974023	47151440	177417	40	11.27287689	9.1327212161	0,002522306
chr3	135917277	136097185	179908	44	5.558396513	17.0092491718	0,00789545
chr6	13138755	13319020	180265	41	5.547388566	12.1532188722	0,004952059
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chrX	134066850	134381777	314927	41	9.526017140	27.3009300568	0,002958487
chr7	120752862	121070772	317910	38	6.291088672	12.2956182567	0,00106783
chr5	39338397	39661094	322697	39	6.197764466	12.3856125096	0,001414362
chr21	35239019	35563219	324200	42	3.084515731	10.4426280074	0,002482301
chr1	196523601	196854892	331291	36	6.036988629	19.6268537328	0,001048088
chr1	179582274	179914924	332650	44	3.006162633	9.0019540057	0,007816023
chr3	100405940	100751363	345423	37	0	32.5111529921	0,000391216
chr15	30981596	31339669	358073	40	2.792726622	12.5041541809	0,002220218
chr5	169607689	169967946	360257	46	8.327388503	7.711161754	0,008140324
chr9	72330731	72691865	361134	40	2.769055253	10.0594239257	0,002526763
chr1	171079626	171441099	361473	38	5.532916704	29.5001286403	0,001317708
chr5	125212437	125580423	367986	37	0	31.8272977776	0,000389263
chr3	125203914	125577201	373287	43	2.678903899	10.2516294433	0,004547907
chr6	71921833	72297005	375172	37	10.66177646	20.2533771177	0,001123809
chr9	88845694	89223374	377680	41	5.295488244	17.0527960178	0,003145003
chr11	129904731	130282681	377950	43	2.645852626	15.380341315	0,006165452
chr1	170367442	170754036	386594	38	5.173386032	18.3844032758	0,001858335
chr3	111574496	111978524	404028	36	0	28.5967804212	0,000886902
chr9	80376445	80786453	410008	38	0	16.0243214767	0,00082155
chr14	71641452	72054996	413544	42	2.418122376	12.5843924709	0,002924777
chr3	29089571	29504969	415398	37	2.407329837	12.0015021738	0,000845819
chr5	35647009	36063024	416015	39	9.615037919	15.2917563069	0,001238861
chr22	24485723	24906652	420929	45	2.375697564	9.8132939284	0,004307932
chr10	3517343	3938745	421402	43	2.373030977	10.6055500449	0,004409423
chr3	24245456	24667486	422030	38	2.369499798	16.6765395825	0,00242562
chr5	169002159	169428113	425954	43	2.347671344	7.829718702	0,003346797
chr5	147947612	148378867	431255	41	6.956441084	18.8883607147	0,002686431
chr7	47191171	47629879	438708	48	2.279420480	10.3380836456	0,00692374
chr2	1212247	1655380	443133	47	6.769976508	12.0677087917	0,007725456
chr2	205896519	206339785	443266	41	4.511963471	9.887967947	0,003950719

chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr10	28034714	28479046	444332		38 6.751708182	19.5074853938	0,001782688
chr15	45995892	46445085	449193		39 8.904858268	10.1332834661	0,002705707
chr6	139797732	140252451	454719		39 0	16.4653335357	0,00139844
chr3	160328834	160790375	461541		38 4.333309500	22.5009262449	0,001128941
chr3	149729634	150191680	462046		37 6.492860018	17.9514593785	0,001795773
chr7	78653349	79119237	465888		35 2.146438629	18.7967923621	0,002078979
chr2	121067557	121538570	471013		51 2.123083651	11.2024508878	0,013829184
chr4	158989114	159463103	473989		37 4.219507203	21.3654747262	0,000997856
chr4	158022726	158500480	477754		36 6.279382276	12.8484533881	0,00095935
chr13	35022669	35503835	481166		40 2.078284833	7.9766650179	0,002680987
chr13	109340305	109832540	492235		44 4.063099942	12.2693428952	0,006247016
chr6	11281465	11790062	508597		41 3.932386545	14.9511302662	0,002599212
chr1	115412781	115937554	524773		42 3.811171687	16.8844433688	0,003375609
chr6	12229125	12756698	527573		39 3.790944570	17.575008577	0,001370572
chr4	166553472	167083145	529673		38 3.775914573	13.2106035233	0,001356348
chr2	149338365	149879950	541585		43 3.692864462	10.2414210142	0,002623651
chr1	241845001	242388266	543265		40 3.681444598	12.1852134778	0,002535595
chr10	117590902	118134526	543624		41 5.518520153	11.1477786117	0,003032943
chr5	146040665	146586320	545655		39 1.832659830	19.2213028379	0,001719881
chr13	108268185	108813915	545730		39 1.832407967	14.1648800689	0,00080344
chr6	165876975	166425534	548559		43 3.645915936	13.5845004822	0,005439198
chr22	25695318	26274047	578729		46 0	8.883259695	0,00769302
chr1	238585693	239166532	580839		40 5.164942436	12.4296061387	0,00111492
chr5	142429949	143019158	589209		40 3.394381280	15.1177256288	0,002214834
chr9	136551724	137142117	590393		53 6.775148079	11.6930586914	0,017583427
chr1	98365114	98963496	598382		36 1.671173263	24.3070480061	0,00086344
chr15	36977859	37586958	609099		39 1.641769236	20.6662627914	0,001363931
chr9	117319871	117935361	615490		38 1.624721766	23.5948593803	0,000679818
chr3	8025135	8669832	644697		41 6.204465043	14.6042249305	0,002443954
chr14	88733047	89388349	655302		44 3.05202792	7.2743559458	0,003450873
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr2	160702875	161378263	675388		39 2.961260786	20.6530172286	0,001457852
chr5	166694257	167378878	684621		39 1.460662176	5.7330990431	0,001535568
chr1	827249	1515403	688154		60 53.76703470	3.9703322222	0,0358568
chr3	112160997	112851431	690434		37 2.896728724	30.2383718067	0,001354416
chr8	119752058	120443579	691521		38 4.338263046	22.5857204626	0,000814376
chr10	53431796	54129533	697737		37 2.866409549	17.6510633663	0,00101099
chr16	84874691	85572680	697989		49 8.596124007	6.9458114669	0,009087331
chr5	82802346	83503229	700883		37 4.280314974	14.0878006743	0,000782796
chr5	32883759	33602044	718285		39 2.784410087	26.9540641946	0,00117802
chr11	28378895	29111820	732925		37 0	17.2501961319	0,000741199
chr1	97566353	98322517	756164		35 2.644928877	13.9461809872	0,000925725
chr1	237788353	238557765	769412		40 2.599387584	11.7948511331	0,001926796
chr7	31589286	32360448	771162		40 3.890233180	20.4881723944	0,002859321
chr2	237081818	237854071	772253		45 2.589824837	12.3117682936	0,005677472
chr20	51492858	52269330	776472		43 7.727258677	9.8344821191	0,002962116
chr2	217616537	218408776	792239		42 2.524490715	14.5471252993	0,004850629
chr3	147127436	147980397	852961		36 5.861932726	22.8332831161	0,000937909
chr9	77880054	78750625	870571		41 5.743356946	9.2860892449	0,002776423
chr3	63688431	64567739	879308		41 5.686289673	10.8128209911	0,003608714
chrX	148491866	149381775	889909		43 10.11339361	27.149629906	0,004771849
chr8	32124590	33042831	918241		38 1.089038716	14.2620510302	0,001372762
chr13	73181416	74121818	940402		38 1.063375024	15.483803735	0,001166914
chr1	208340376	209293324	952948		42 4.197500808	19.7026490428	0,003770255
chr1	216193837	217171537	977700		39 2.045617265	14.1865602946	0,001201145
chr21	26941307	27938615	997308		37 2.005398532	14.4659423167	0,000804869
chr2	204432865	205431940	999075		39 3.002777569	14.5311413057	0,001714406
chr2	226921801	227941946	1020145		40 5.881516843	12.6553578168	0,003055938
chr8	18603905	19655183	1051278		41 2.853669533	12.8373275195	0,002647185
chr13	42500654	43591236	1090582		39 3.667766385	17.0331070933	0,002311163
chr10	62858856	64101326	1242470		39 4.024242034	15.7947475593	0,002121498
chromosome	start	end	size in bp	% in GC	Genes/Mb	% in line1	G4/GC
chr18	54702074	55998099	1296025		42 6.944310487	15.0969310006	0,00276119
chr20	8316866	9718874	1402008		39 2.853050767	10.3632076279	0,001309477
chr8	96970949	98490242	1519293		40 4.607406208	18.5423088239	0,001995994
chrX	9804380	13178880	3374500		40 3.852422581	14.6900281523	0,002615943

III. Evidences supplémentaires du rôle de Polθ dans le *timing* de réplication

Cette étude a donc permis de mettre en évidence un rôle inédit pour une ADN polymérase, la régulation du *timing* de réplication, et ce par un recrutement à la chromatine très tôt en phase G1, de façon concomitante avec le TDP, au niveau de complexes de pré-réplication. Le fait que Polθ joue un rôle en G1 dans la régulation temporelle de l'activation des origines de réplication en phase S est étayé par des expériences supplémentaires non publiées.

A. Polθ joue un rôle dans l'activation des origines de réplication

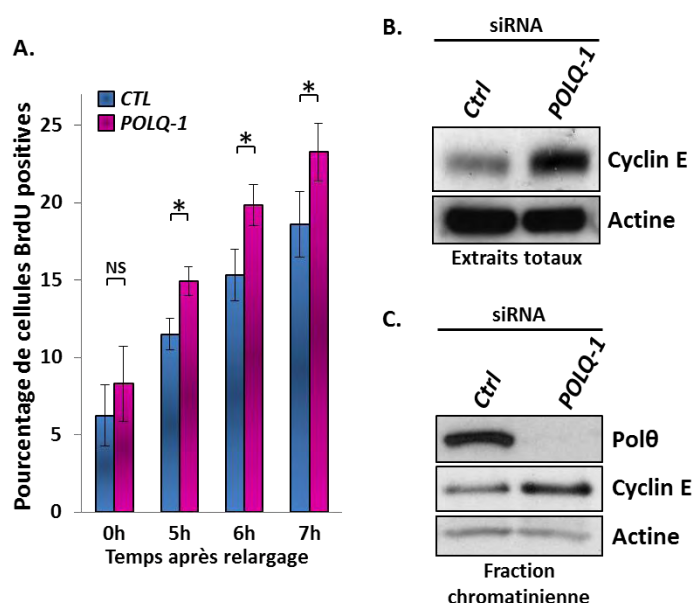


Figure 18 : La déplétion de Polθ induit une activation précoce des origines de réplication. (A) 24h après transfection par les siRNA contrôle (CTL) ou dirigé contre *POLQ* (POLQ-1), les cellules ont été synchronisée en mitose puis relarguées dans le cycle. Au différents temps de récolte, une incorporation de BrdU a été réalisée et les cellules positives quantifiées par cytométrie en flux. L'expérience a été répétée trois fois et un test de Student a été appliqué (* $p < 0.01$) (B) 24h après transfection par les siRNA CTL ou POLQ-1, les cellules ont été récoltées et la quantité de cycline E dans les extraits totaux ont été analysés par western blot. (C). 24h après transfection par les siRNA CTL ou POLQ-1, les cellules ont été synchronisées en mitose par traitement au nocodazole puis relarguées dans le cycle pendant 6h pour être récoltée. Après fractionnement subcellulaire, la fraction chromatiniennne a été analysée par Western Blot. Les expériences (B) et (C) ont été réalisées au moins trois fois.

Parmi les premières expériences réalisées par Anne Vidal-Fernandez au laboratoire pour évaluer le rôle de Polθ dans la réplication, l'analyse du cycle cellulaire par cytométrie en flux a révélé une entrée des cellules en phase S précoce dans les cellules déplétées pour *POLQ*. En effet, après synchronisation des cellules en mitose par un traitement au nocodazole, le pourcentage des cellules positives pour le BrdU après différents temps de relargage est légèrement mais significativement plus élevé dans les cellules déplétées pour *POLQ* (Figure

18A). L'augmentation du pourcentage de cellules ayant incorporé du BrdU, et donc étant entrées en phase S, suggère que, en absence de Pol θ , les cellules sont rentrées plus tôt en phase S.

D'autre part, l'analyse par western blot d'extraits totaux sur cellules asynchrones (Figure 18B) ou à la chromatine à la transition G1/S (Figure 18C) montrent une augmentation du taux cellulaire de cycline E, et un recrutement accru à la chromatine. L'ensemble de ces résultats confirme le lien entre Pol θ et l'activation des origines de réplication, certainement par la modulation de l'expression de la Cycline E.

B. Pol θ influence l'environnement et l'organisation chromatinien

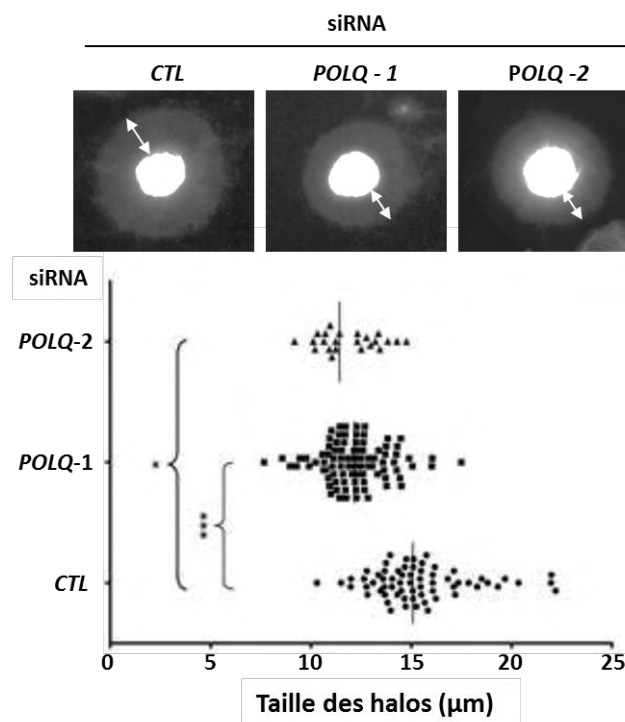


Figure 19 : La déplétion de Pol θ modifie l'environnement chromatinien. 30h après transfection des cellules RKO par les siRNA *CTL* ou dirigés contre *POLQ* (*POLQ-1* et *POLQ-2*), les cellules ont été synchronisées en mitose par traitement nocodazole. Après relargage, elles ont été synchronisées à la transition G1/S par un traitement à la L-mimosine. Les noyaux ont été traités pour former les Halos d'ADN selon le matériel et méthode. Leur rayon a été mesuré. (*CTRL* siRNA: n=65, *POLQ-1* siRNA : n=90, *POLQ-2* siRNA : n=23). * : $p < 0,05$; ***: $p < 0.001$.

Nous nous sommes aussi demandé si la modification du *timing* de réplication pouvait être due à une réorganisation spatiale des origines de réplication. En effet, comme nous l'avons vu en introduction, les origines de réplication sont regroupées dans des foyers de réplication au sein desquels l'activation des origines est synchrone. Cette organisation spatiale implique la formation des boucles d'ADN. Nous avons donc cherché à savoir si l'organisation nucléaire était modifiée en absence de Pol θ . Pour cela, ont été menées des expériences de

visualisation par fluorescence des Halo d'ADN permettent de visualiser ces boucles de chromatine décondensées. Ces expériences, réalisées par Anne Vidal-Fernandez, ont montré qu'en absence de Polθ, à la transition G1/S, la taille des halos diminue (Figure 19). L'expression de Polθ influence donc l'organisation des boucles de chromatine avant l'entrée en phase S, certainement au moment de la formation des noyaux interphasiques en début de phase G1. Ainsi, Polθ joue donc potentiellement un rôle dans le choix des origines de réplication activées au cours d'une phase S.

C. Polθ régule le *timing* de réplication dans des cellules humaines normales.

Dans l'article, les expériences de *timing* de réplication ont été menées sur une lignée cellulaire tumorale, les RKO, ou une lignée transformée par le SV40, les MRC5-SV. Pour vérifier que le rôle de Polθ dans la régulation du *timing* de réplication soit un processus physiologique normal, qui ne soit pas restreint aux cellules transformées, nous avons réitéré des expériences de *timing* de réplication sur des cellules normales, les MRC5-N. L'hybridation sur puce des fractions « début de phase S » dans les conditions contrôle ou après extinction de *POLQ* ont montré des modifications de *timing* dans ce troisième modèle cellulaire (Figure 20), confirmant le rôle général de cette ADN polymérase dans la régulation temporelle des origines de réplication, quel que soit le type cellulaire.

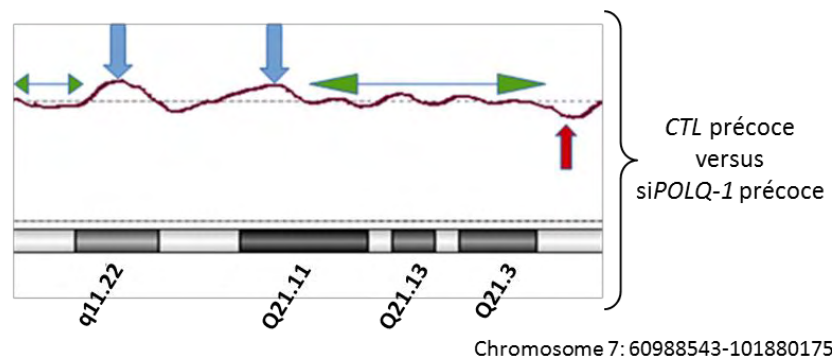


Figure 20 : Le *timing* de réplication des MRC5-N est modifié en absence de Polθ. 36h après transfection par les siRNA *CTL* ou *POLQ-1*, une expérience de *timing* de réplication a été réalisée. Les fractions précoces ont été hybridées sur puce et un exemple des résultats obtenus pour le chromosome 7 est présenté. Les domaines où le *timing* n'est pas modifié sont indiqués par la double flèche verte, les domaines retardés après extinction de *POLQ* par une flèche bleue, et les domaines avancés par une flèche rouge.

IV. Conclusion et Perspectives

A. Conclusion

L'ensemble de ces travaux expérimentaux publiés et non publiés montrent que l'ADN polymérase Thêta joue un rôle dans la réplication sans stress exogène. Ces résultats apportent

un nouvel argument en faveur d'une reconsidération du rôle de ces ADN polymérases translésionnelles. En effet, de plus en plus d'évidences montrent qu'elles interviennent aussi dans la réplication génomique en absence de stress exogène. Mon équipe d'accueil l'a par exemple montré pour deux d'entre elles : Pol η et Pol κ . La première participe à la réplication des sites fragiles (Bergoglio et al., 2013; Bétous et al., 2009) et des résultats non publiés suggèrent que ce soit aussi le cas pour d'autres séquences particulières du génome. Pol κ est, elle, nécessaire au bon fonctionnement du point de contrôle de phase S (Bétous et al., 2013).

B. Perspectives

1. Mécanismes d'action de Pol θ dans sa fonction régulatrice du *timing* de réplication.

Dans la continuité de ces travaux, il sera intéressant de comprendre comment Pol θ exerce son rôle dans la régulation du *timing* de réplication et de préciser son mécanisme d'action. Les résultats obtenus montrent une modulation du *timing* de réplication sur 5 à 20% du génome, après extinction ou surexpression respectivement de l'ADN polymérase Théta. Parallèlement, les expériences de co-immunoprécipitation suggèrent que Pol θ est localisée au niveau d'origines de réplication. Il sera donc intéressant de déterminer quelles sont les cibles de Pol θ , c'est-à-dire sur quelles séquences elle vient se fixer en phase G1. En particulier, Pol θ comportant un domaine hélicase, et les origines de réplication ayant été associées à la présence des structures secondaires d'ADN, les G-quadruplex, on peut spéculer sur un type de régulation via la déstructuration de ces G4. Pour répondre à cette question des séquences de recrutement de Pol θ à la chromatine, nous avons envisagé une approche par immunoprécipitation de la chromatine suivie d'un séquençage des fragments d'ADN immunoprécipités (ChIP-seq). Au cours de ma quatrième année de thèse, j'ai mis au point les conditions de fragmentation de l'ADN par digestion enzymatique et d'immunoprécipitation de Pol θ . J'ai réalisé les ChIP sur deux lignées cellulaires : les cellules RKO synchronisées en G1, dans lesquelles les précédentes études ont été faites, et les K562, cellules pour lesquelles nous disposons de nombreuses données génétiques et épigénétiques accessibles. Les expériences sont en cours de finalisation et les échantillons devraient être envoyés pour séquençage dans les semaines à venir.

Le recrutement d'une ADN polymérase en phase G1 du cycle cellulaire paraît contre-intuitive étant donné qu'il n'y a priori pas de synthèse d'ADN au cours de cette phase du cycle cellulaire. Cependant, il est possible que le recrutement ait lieu en G1 mais son activation au cours de la phase S suivante. De plus, comme évoqué précédemment, Pol θ

possède aussi un domaine hélicase qui peut être impliqué dans cette fonction. Pour déterminer si un des domaines en particulier est important pour son recrutement en G1 et son rôle dans le *timing* de réplication, il sera intéressant de travailler avec des mutants de la protéine pour évaluer leur impact sur le recrutement de Polθ et des Mcm à la chromatine en G1, ainsi que sur le *timing* de réplication. Comme décrit précédemment (Yousefzadeh et al., 2014), l'inactivation du domaine polymérase provient de mutations de 2 acides aminés D2330A et Y2331A. L'inactivation du domaine hélicase est issue de la substitution dans le motif hélicase I, de la lysine K120 par une alanine (K121A). En effet, la mutation de ce résidu dans HEL308, homologue du domaine hélicase de Polθ abolit son activité hélicase. Deux stratégies peuvent être envisagées ensuite : l'expression transitoire des mutants après extinction de la protéine endogène par un siRNA ciblant la région non codante 3'UTR de l'ARNm de *POLQ*, ou l'utilisation du système CRISPR/Cas9 qui permet de muter directement la protéine endogène *in situ*. Ce système a le double avantage de s'affranchir totalement de la forme sauvage de la protéine, et de conserver une régulation physiologique de l'expression des mutants de Polθ. En revanche, il implique de travailler avec des mutants stables qui peuvent mettre en place des voies de compensation et d'adaptation à la mutation.

Enfin, une dernière approche pour comprendre les mécanismes de régulation de Polθ sur le *timing* de réplication sera d'en identifier les partenaires en couplant les IP à des analyses en spectrométrie de masse. De façon à concentrer nos recherches sur son rôle dans le *timing* de réplication, et non dans la réparation ou l'élongation en cours de phase S, ces IP seront réalisées en phase G1, qui est je le rappelle la phase où se décide le *timing* de réplication.

2. L'impact de la modification du *timing* de réplication suite à la dérégulation de Polθ

Si nous avons observé une modification du *timing* de réplication en fonction du statut de Polθ, les conséquences cellulaires de ces modifications sont totalement inconnues. En particulier, il serait intéressant d'évaluer si des modifications transcriptionnelles, ou en terme d'instabilité génétique ont lieu suite à cette dérégulation temporelle de l'activité des origines.

Tout d'abord, comme nous l'avons vu en introduction, il existe un lien entre transcription et *timing* de réplication, les régions précoces étant majoritairement transcriptionnellement actives, et les régions tardives réprimées. Il sera donc intéressant de voir si ces modifications du *timing* de réplication s'accompagnent d'une modification de l'expression génique en fonction du statut de Polθ, et si tel est le cas, si un processus biologique est en particulier affecté. En absence de Polθ, les gènes contenus dans les

domaines dont le *timing* est modifié sont impliqués dans une multitude de voies de signalisation (Annexe1). Seule l'analyse transcriptionnelle pourra donc dire si l'expression des gènes est modifiée pour une ou plusieurs voies en particulier. Dans le cas où Polθ est surexprimée, le retard de réplication observé suggère qu'un certain nombre de gènes, contenus dans les régions affectées, pourrait être réprimés. Or, compte tenu de la signature de cette ADN polymérase dans les cancers, il serait intéressant de voir si des gènes suppresseurs de tumeurs peuvent être par exemple réprimés, comme cela a été décrit dans le cas du cancer du sein (Fritz et al., 2013).

Ensuite, les régions répliquées en fin de phase S sont soumises à une instabilité génétique plus forte que les régions précoces (mutations ponctuelles, amplifications, translocations). Et les travaux de Lemée *et al.* ont montré que la surexpression de Polθ induit une forte instabilité chromosomique (Lemée et al., 2010). Des analyses cytogénétiques ciblant les régions dont le *timing* est modifié pourraient permettre de voir si de telles modifications génétiques sont particulièrement enrichies dans ces régions retardées.

CHAPITRE II : MODIFICATION DU *TIMING* DE RÉPLICATION INDUITE PAR UN STRESS RÉPLICATIF AU CYCLE PRÉCÉDENT

I. La transmission des dommages suite à un stress réplicatif

A. Le stress réplicatif

1. Définition et causes du stress réplicatif

La progression des fourches de réplication est en permanence mise à l'épreuve par des obstacles ou situations qui vont ralentir ou empêcher la progression des ADN polymérases répliquatives (Figure 21) et qui, si la situation persiste, peuvent engendrer un effondrement de la fourche. L'accumulation de ces événements est regroupée sous le nom de « stress réplicatif » (Gelot et al., 2015; Zeman and Cimprich, 2014). Un stress réplicatif est donc une source d'instabilité génétique et constitue une caractéristique intrinsèque des cellules pré-néoplasiques et cancéreuses, participant à la tumorigenèse (Bartkova et al., 2005; Gorgoulis et al., 2005; Halazonetis et al., 2008).

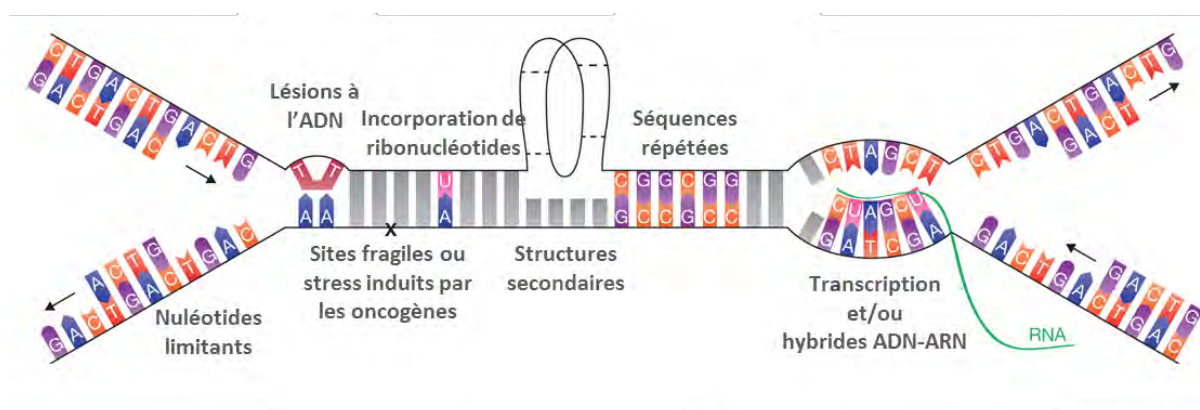


Figure 21 : Les sources de stress réplicatif (Tiré de (Zeman and Cimprich, 2014))

a. Les dommages à l'ADN

Une des plus importantes sources du SR est constituée par les dommages à l'ADN. Ces dommages peuvent être d'origine endogène, inhérents au métabolisme de la cellule, ou exogène, induits par des agents chimiques ou physiques génotoxiques (pour revue (Ciccia and Elledge, 2010)). Parmi les dommages endogènes, on compte les sites abasiques, spontanés ou intermédiaires de la réparation des lésions à l'ADN, les lésions induites par le stress oxydatif dont la plus abondante est la 8-oxo-guanine, la déamination de la cytosine à l'origine de la présence anormale d'uracile dans l'ADN, qui peut aussi être induite par une mauvaise incorporation de base, ou la méthylation de bases. Les dommages de source exogène sont très

nombreux et sont liés à l'environnement de l'organisme. Les plus étudiés sont les dimères de pyrimidine induits par les rayons ultra-violet, les adduits aromatiques causés par de puissants carcinogènes de l'environnement, ou les cassures double-brins dues aux radiations ionisantes. Cependant les dommages à l'ADN ne constituent pas la seule source de stress réplicatif.

b. La dérégulation des facteurs de réplication

La réplication est le résultat d'un équilibre entre activation des origines et cinétique d'élongation. Ces deux mécanismes font appels à de nombreux facteurs, dont certains sont limitants. Il est donc important que ces facteurs soient présents en quantité suffisante pour répondre à ces deux composantes. Tout d'abord, la disponibilité en deoxynucléotides (dNTP) doit être finement régulée, quantitativement, mais aussi qualitativement. En effet, une déficience en dNTP ralentit la progression des fourches de réplication, induisant l'activation d'origines de réplication dormantes. Si cette déficience est trop importante, la progression des fourches de réplication est perturbée, induisant de l'ADN simple-brin et l'activation des réponses aux dommages à l'ADN. De cette façon, une dérégulation du pool de dNTP induit de l'instabilité génétique et participe aux stades précoces de la tumorigenèse (Anglana et al., 2003; Bester et al., 2011; Poli et al., 2012). La déplétion en dNTP peut être induite par inhibition de la ribonucléotide réductase (RNR) ou la Cytidine Déaminase. Elle peut également venir de l'activation d'un trop grand nombre d'origines de réplication, liée à la surexpression de la cycline E ou d'une suractivation de la voie Rb-E2F qui régit l'entrée en phase S (Beck et al., 2012). Par ailleurs, un déséquilibre dans la composition du pool de dNTP peut également constituer une source de stress réplicatif. Par exemple, une surreprésentation de la déoxyCytidine-Nucléo-triphosphate (dCTP) induite par la dérégulation de la Cytidine Déaminase, induit une augmentation de l'instabilité génétique observée dans le syndrome de Bloom, ainsi que l'apparition de régions sous-répliquées qui persistent en mitose (Chabosseau et al., 2011; Gemble et al., 2015). Par ailleurs, de faibles doses d'hydroxyurée, un inhibiteur de la RNR, induisent un stress réplicatif léger à l'origine d'une amplification des centrosomes en mitose conduisant à une mauvaise ségrégation des chromosomes, générant ainsi de l'instabilité chromosomique (Wilhelm et Magdalou et al., 2014).

Deuxièmement, la régulation de l'activation des origines de réplication est importante pour la stabilité du génome. Comme nous l'avons vu en introduction, cette activation finement régulée dépend de facteurs limitants tels que Cdc45. D'un côté, si ces facteurs ne sont pas suffisants, la cellule peut souffrir d'un défaut d'activation d'origines qui peut avoir des conséquences à plusieurs niveaux. Tout d'abord, cela peut empêcher l'activation

d'origines dormantes en cas de problème de réplication. Cela peut aussi engendrer des zones sous répliquées en fin de phase S dans des régions où la densité de sites d'initiation est trop faible. Inversement, une activation d'un nombre trop important d'origines de réplication est aussi source de stress réplcatif et d'instabilité génétique. En effet, cela entraîne un épuisement des ressources en dNTP, mais aussi en RPA, recruté au niveau des très nombreuses zones d'ADN simple brin, et qui fera défaut pour la protection d'une population d'ADN simple brin en cas de découplage entre hélicase et polymérases qui constitue un site privilégié de cassures de l'ADN (Toledo et al., 2013).

Enfin, la duplication du génome s'accompagne de la formation de nouveaux nucléosomes au niveau du brin néosynthétisé et du brin parental, exigeant une quantité importante d'histone. Ainsi, un déficit en nouvelles histones ou un problème dans l'assemblage du nucléosome perturbe la progression des fourches de réplication, induisant s'ils persistent, un effondrement du réplisome et des cassures chromosomiques (Mejlvang et al., 2014).

c. La séquence nucléotidique

Le stress réplcatif peut être induit par la séquence même de l'ADN : les séquences nucléotidiques répétées. Elles représentent 50% du génome humain et peuvent former des structures secondaires intramoléculaires de l'ADN, appelées ADN non-B, par opposition à la conformation physiologique en B de la double-hélice d'ADN (pour revue (Boyer et al., 2013)). La formation de ces structures a lieu lorsque l'ADN est sous-forme simple-brin, donc pendant la réplication ou la transcription. Parmi ces séquences répétées, on distingue plusieurs catégories : répétitions inversées, répétitions en miroir ou répétitions directes. Parmi les répétitions directes, on trouve les séquences microsatellites, minisatellites ou satellites, en fonction du nombre de bases répétées (respectivement 1 à 10nt, 10 à 60nt et plus de 60nt). Les séquences satellites sont en particulier retrouvées au niveau des centromères et télomères. En fonction de la composition en nucléotides, les séquences répétées peuvent ainsi donner lieu à différents types de structures telles les épingles à cheveux, cruciformes, triple-hélices, ADN-Z ou G-quadruplex. Ces derniers sont formés dans des régions riches en G, dont la séquence n'est pas précisément définie mais répond néanmoins au motif suivant : $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$. Ils sont retrouvés aux télomères, mais aussi dans les promoteurs ou les gènes ribosomaux (Huppert, 2010). Les régions riches en AT constituent des régions flexibles capables de former structures secondaires de type épingles à cheveux (Zlotorynski et al., 2003). Elles sont en particulier retrouvées dans des régions particulières du génome appelées

sites fragiles communs que nous détaillerons plus tard. Ces structures peuvent avoir un rôle physiologique important, à l'image des G4 qui sont associés aux origines de réplication et seraient impliqués lors de la transcription. Cependant, elles constituent aussi des barrières de fourche de réplication car les ADN polymérases réplcatives sont incapables de les franchir aisément. Elles nécessitent donc l'intervention d'hélicases ou d'ADN polymérases spécialisées permettant leur réplication, rendant possible le redémarrage des fourches. Ces séquences répétées sont aussi des points chauds de décalages de bases suite à des glissements des extrémités 3'OH qui peuvent s'hybrider en amont ou en aval de leur position normale, induisant des rétractions ou expansions des répétitions (Kim and Mirkin, 2013).

L'ensemble de ces événements peuvent conduire au développement de différentes maladies dont le cancer. L'instabilité des séquences microsatellites est par exemple associée à un grand nombre de tumeurs ainsi qu'au syndrome de Lynch (Imai and Yamamoto, 2008).

d. Les conflits entre réplication et transcription

Les processus de transcription et de réplication, qui partagent le même substrat, sont forcément amenés à interférer. Ces interférences entre réplication et transcription ont été observées au niveau de gènes hautement transcrits, ou dont la transcription peut durer sur plus d'un cycle cellulaire quand le gène est très long. La surexpression de la cycline E, en induisant une activation surnuméraire des origines de réplication, augmente la probabilité de ces collisions entre les deux machineries cellulaires (Jones et al., 2013). Les mécanismes par lesquels la transcription perturbe la réplication sont multiples. D'une part, cette rencontre peut aboutir à des conflits topologiques induisant des nœuds de la double-hélice à l'interface entre les fourches de réplication et l'élongation de la transcription (Olavarrieta et al., 2002). D'autre part, au cours de la transcription, l'hybridation de l'ARN néosynthétisé avec l'ADN qui sert de matrice à la transcription peut induire la formation de structures tri-caténaires relativement stables appelées les *R-loops*, qui représentent des obstacles à la progression des fourches de réplication et peuvent générer des CDB (Sollier and Cimprich, 2015; Tuduri et al., 2009). En effet, l'inhibition de facteurs limitant la formation des *R-loops*, ou impliqués dans leur résolution, stimule des CDB liées à la réplication. Et inversement, l'inhibition de la réplication induit une augmentation des dommages au niveau de ces structures (Alzu et al., 2012; Gan et al., 2011; Sollier et al., 2014; Tuduri et al., 2009; Yüce and West, 2013). La RNA :DNA hélicase senataxine par exemple voyage avec la fourche de réplication pour favoriser le franchissement des *R-loop* au niveau de gènes transcrits par l'ARN polymérase II, limitant une accumulation de dommages en phase S-G2 (Alzu et al., 2012; Yüce and West,

2013). D'autre part, la topoisomérase I (TopoI) inhiberait la formation des *R-loops*, limitant ainsi le blocage et l'effondrement des fourches de réplication à leur niveau. Ce rôle de TopoI dans la stabilité génétique en réponse aux *R-loops* est dépendante des facteurs ASF1 et 2 (pour « *Anti-silencing function protein 1 and 2* ») (Tuduri et al., 2009). Il est intéressant de noter que les cassures induites par l'inhibition d'ASF1 sont diminuées par l'inhibition de XPG. En effet, il a été montré que les nucléases XPG et XPF sont responsables de cassures de l'ADN au niveau des *R-loops*. Leur réparation implique des facteurs impliqués dans une voie de réparation couplée à la transcription, le TC-NER (pour « *Transcription Coupled Nucleotid Excision Repair* ») (Sollier et al., 2014). Il se peut donc que pendant la réplication, les nucléases XPG et XPF soient nécessaires pour l'éviction des *R-loops* qui sont ensuite réparées par le TC-NER.

Ainsi, bien que les mécanismes ne soient pas clairement élucidés, la formation des *R-loops* au cours de la transcription représente une source majeure du conflit entre transcription et réplication pouvant générer de l'instabilité génétique.

2. Le cas particulier des sites fragiles communs

Les sites fragiles communs (SFC) sont de longues séquences qui réunissent à elles seules différentes sources de stress réplicatif. Elles sont flexibles, riches en AT, et sont définies par leur aptitude à casser en condition de stress réplicatif, ce qui conduit fréquemment à des réarrangements chromosomiques. Les raisons de leur fragilité sont multiples et sont certainement le résultat d'une combinaison de facteurs (Debatisse et al., 2012; Durkin and Glover, 2007). Tout d'abord, les séquences non-B qu'ils contiennent sont propices à la formation de structures secondaires. De plus, ces régions, situées dans les régions de transition du *timing* de réplication, sont répliquées en fin de phase S et leur réplication ne s'achève souvent qu'en phase G2 du cycle cellulaire (Achkar and Gerbault-Seureau, 2005). Ce retard de réplication s'explique en particulier par une faible densité en origines de réplication, les fourches de réplication ayant à synthétiser de longs domaines (Letessier et al., 2011; Le Tallec et al., 2012). Si un obstacle rencontré persiste, comme une structure secondaire par exemple, l'absence d'origines dormantes de secours participe à leur instabilité. En outre, les SFC abritent de longs gènes, et leur transcription peut être corrélée à la fragilité de la séquence du fait de collisions entre les machineries de réplication et de transcription (Helmrich et al., 2006, 2011). Enfin, l'environnement chromatinien influence certainement aussi la fragilité des SFC, que ce soit en termes d'organisation de la chromatine ou de modifications d'histones.

3. Réponse cellulaire au stress réplcatif

La progression de la fourche est donc en permanence mise à l'épreuve par des dommages d'origine endogène ou exogène. Au-delà d'un certain seuil de perturbations, la réponse aux dommages à l'ADN est engagée et une voie de réponse majeure au stress réplcatif est activée, le point de contrôle de phase S jouant un rôle crucial dans le maintien de l'intégrité du génome (Bartek et al., 2007; Zhou and Elledge, 2000). Il consiste en une cascade de signalisation dont la kinase ATR en est le principal acteur (Figure 22). Le découplage entre hélicases réplcatives et ADN polymérase réplcative bloquée induit l'apparition d'ADN simple brin qui est recouvert par RPA. Par l'intermédiaire d'ATRIP (pour « *ATR Interacting Protein* »), l'accumulation de RPA entraîne le recrutement de la sérine/thréonine kinase ATR (pour « *ATM and Rad3 Related* »). Son activation nécessitera d'autres événements, tels que le recrutement du complexe 9-1-1, trimère composé de Rad9, Hus1 et Rad1, de TopBP1 (pour « *Topoisomerase Binding Protein I* »). Enfin, la synthèse d'amorces d'ARN :ADN à la fourche découplée, par la primase Pol α d'abord puis les ADN polymérases δ et κ est nécessaire à une pleine activation d'ATR (Bétous et al., 2013; Van et al., 2010). ATR activé va alors phosphoryler un de ses principaux effecteurs la sérine/thréonine kinase Chk1 sur les sérines 345 ou 317. Cette phosphorylation de Chk1 par ATR est dépendante de la Claspin, qui pallie à la faible affinité d'ATR pour Chk1 (Liu et al., 2006; Mailand et al., 2006). ATR et Chk1, en phosphorylant leur cibles, vont induire une stabilisation des fourches de réplcation, parallèlement à un arrêt du cycle cellulaire et une inhibition de l'activation des origines tardives (Nam and Cortez, 2011). En revanche, comme nous l'avons vu précédemment, les origines dormantes adjacentes peuvent être activées. Le fait de figer ainsi le système permet la gestion du dommage par la cellule, la réplcation pouvant alors recommencer. Cependant, si le problème persiste, le point de contrôle de phase S peut aussi induire la sénescence ou l'apoptose, protégeant ainsi l'organisme d'une instabilité génétique importante.

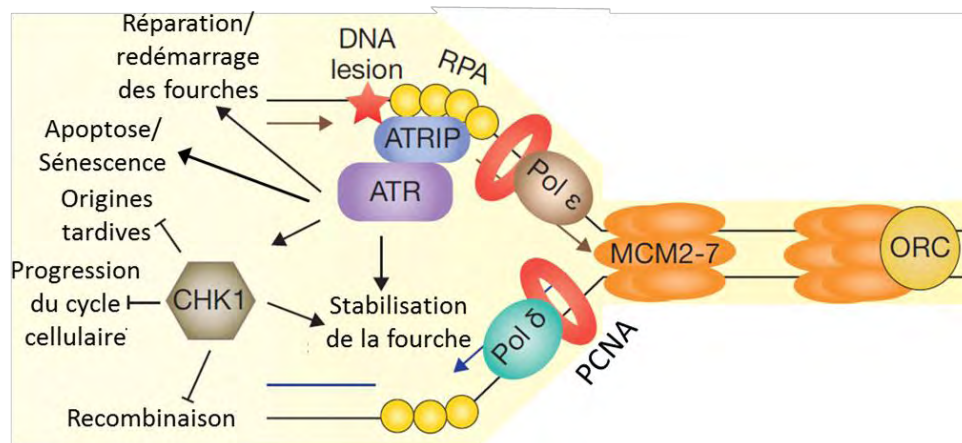


Figure 22 : Point de contrôle de phase S en réponse à un stress réplcatif. Tiré de (Zeman and Cimprich, 2014). Lorsque les ADN polymérases réplcatives sont bloquées par un obstacle, le point de contrôle de phase S est activé : La fourche bloquée est stabilisée, le cycle cellulaire arrêté de façon à permettre la gestion de cet obstacle et le redémarrage de la fourche. Si le blocage persiste, la cellule peut entrer en apoptose ou sénescence.

B. La transmission à la génération suivante de dommages induits par le stress réplcatif

En cas de stress réplcatif ou de dommages à l'ADN importants, l'activation du point de contrôle de phase S empêche donc le passage de cellules endommagées en mitose, dans le but d'éviter une instabilité génétique en mitose. Cependant, un stress réplcatif léger peut échapper à ce point de contrôle de phase S. La kinase ATR est recrutée, mais la cascade de signalisation n'est pas déclenchée (Koundrioukoff et al., 2013). Cela permet alors aux cellules de rentrer en mitose alors que des problèmes liés à la réplication subsistent : ADN sous-répliqué, structures d'ADN aberrantes. Au niveau des sites fragiles, en fin de phase G2, début de mitose, les nucléases Mus81/EME1 et ERCC1 participent à la résolution de ces structures, permettant une bonne ségrégation des chromosomes en mitose (Naim et al., 2013; Ying et al., 2013). Cependant, dans le cas où ces problèmes persistent, ils engendrent en anaphase la formation de structures particulières, les « *DNA UltraFine Bridges (UFB)* » (littéralement les ponts d'ADN ultrafins) (Pour revue (Liu et al., 2014)). Ce sont des séquences d'ADN simple-brin qui relient les deux chromatides sœur lors de leur séparation pendant la mitose. Cette liaison peut avoir lieu entre les centromères, les télomères ou au sein des bras chromosomiques et correspondent alors à des sites fragiles sous-répliqués. Toutes sont recouvertes par l'hélicase Bloom, en association avec la topoisomérase III α , les protéines RMI1 et 2 (pour « *RecQ Mediated genome Instabilty* ») et l'hélicase PICH (Pour « *Plk1-interacting Checkpoint Helicase* »). En revanche, certaines protéines associées sont spécifiques de la localisation des UFB. En particuliers, les UFB situés dans les bras

chromosomiques sont aussi recouverts par RPA et marqués à leurs extrémités par les protéines FANCD2-FANCI (Figure 23). La disparition des UFB au fur et à mesure de l'anaphase suggère que des mécanismes existent dans la cellule pour résoudre ces structures. En effet, sans que les mécanismes moléculaires ne soient connus, l'absence de Bloom ou FANCD2 induit une accumulation de ces UFB en mitose (Chan et al., 2007, 2009; Naim and Rosselli, 2009a). Récemment Rif1 a été aussi identifié comme un acteur participant à la résolution de ces structures intermédiaires (Hengeveld et al., 2015). Enfin, si ces UFB ne sont pas résolus et persistent en télophase, l'ADN finit par casser au cours de la télophase, induisant des dommages qui seront transmis à la cellule fille, et protégé au sein de gros foyers 53BP1, protéine de la réparation des CDB, pendant la phase G1 (Figure 23). Cette transmission de dommages à l'ADN d'une génération à l'autre a été mise en évidence il y a quelques années (Harrigan et al., 2011; Lukas et al., 2011). Il est intéressant de noter que ces foyers sont générés spontanément dans la cellule, et augmentent en cas de stress réplcatif, même léger, appliqué au cycle précédent. En outre, les SFC sont enrichis au niveau de ces structures et un défaut de réplication des SFC induit une augmentation des dommages transmis (Bergoglio et al., 2013; Lukas et al., 2011).

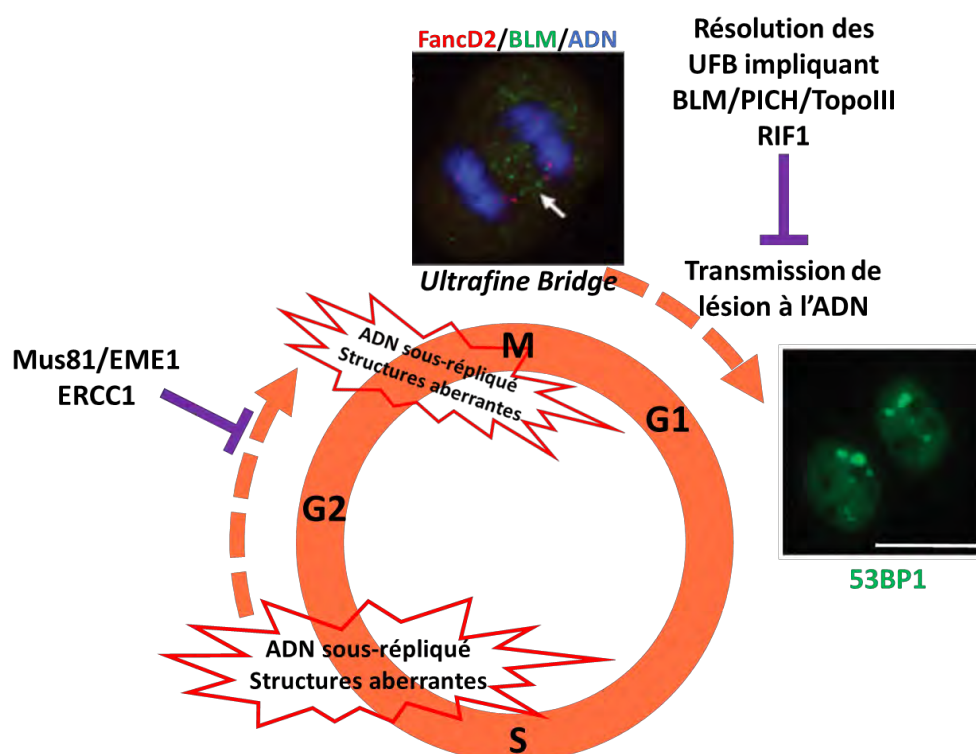


Figure 23 : Transmission de dommages à l'ADN à la génération suivante. Au cours de la phase S, des régions sous-répliquées ou structures en caténane peuvent être générées. Elles peuvent être résolues en G2/M par les endonucléases Mus81-EME1 ou ERCC1. Sinon, elles induisent la formation d'UFB en mitose qui, s'ils ne sont pas pris en charge, cassent et génèrent des CDB protégées par les foyers 53BP1 en phase G1 du cycle d'après. Photos tirées de (Lukas et al., 2011; Naim and Rosselli, 2009b)

C. 53BP1 et réparation des cassures double brin

Le devenir de ces dommages transmis est pour le moment non élucidé. Ils sont protégés par les foyers 53BP1 pendant la phase G1 et à l'entrée en phase S, ces foyers ont disparu (Lukas et al., 2011). Deux hypothèses sont possibles. Ils peuvent être d'une part protégés dans les foyers 53BP1 durant toute la G1 pour leur permettre de la traverser jusqu'à la phase S où leur réplication sera alors terminée, ou leur réparation assurée. D'autre part, on peut envisager que leur réparation a lieu au cours de la phase G1, au sein de ces foyers. En effet, 53BP1 est un acteur connu de la réparation des CDB, qui détermine la voie de réparation que la cellule doit engager.

1. Le recrutement de 53BP1 aux CDB

Le recrutement de 53BP1 au niveau des dommages est régulé de plusieurs façons.

Tout d'abord, deux modifications d'histones induisent principalement le recrutement de 53BP1 au niveau de la lésion. Premièrement, la diméthylation de l'histone H4K20 (H4K20me2) est nécessaire à la formation des *foci* 53BP1 après dommages à l'ADN, 53BP1 interagissant directement avec elle au niveau de son domaine Tudor (Botuyan et al., 2006). Cependant, cette marque d'histone est ubiquitaire et non spécifique au CDB. Il existe donc des signaux supplémentaires qui permettent à 53BP1 de cibler spécifiquement les sites de dommages à l'ADN (pour revue (Panier and Boulton, 2014)). Tout d'abord, l'ubiquitination par les ubiquitine-ligases RNF8-RNF168 de deux protéines en compétition avec 53BP1 pour sa liaison avec H4K20me2 (Polycombe L3MBTL1 et la protéine déméthylase 2A contenant un domaine Jumonji (JMJD2A)) permettrait leur éviction, rendant l'histone H4K20me2 accessible à 53BP1. De plus, la déacétylation de l'histone H4 sur sa lysine 16 par les Histones Déacétylases HDAC1 et 2 est aussi nécessaire car cette modification de l'histone H4 empêche l'interaction de 53BP1 avec H4K20me2. D'autre part, l'ubiquitination de l'histone H2A sur sa lysine 15 (H2AK15Ub) s'est révélée indispensable au recrutement de 53BP1 au niveau des dommages à l'ADN (Fradet-Turcotte et al., 2013). Cette ubiquitination est catalysée par RNF168, spécifiquement au site de la CDB, ciblant le recrutement de 53BP1 à son niveau, via son interaction avec H2AK15Ub au niveau de son motif de recrutement dépendant de l'ubiquitine (motif UDR).

Deuxièmement, le recrutement de 53BP1 aux CDB est régulé par sa phosphorylation sur deux sites situés dans le domaine UDR (T1609 et S1618). En effet, d'un côté la phosphorylation de 53BP1 par les kinases Cdk1 et Plk1 en mitose inhibe son recrutement à la chromatine pendant cette phase du cycle, phase pendant laquelle la réparation des dommages

à l'ADN se révèle nocive pour la cellule (Orthwein et al., 2014). D'un autre côté, le maintien de cette phosphorylation par l'inhibition de la phosphatase PP4C ou de sa sous-unité régulatrice PP4R3 β , empêche le recrutement de 53BP1 au site des dommages à l'ADN, en particulier en phase G1, quand les cellules sortent de mitose (Lee et al., 2014).

Enfin, le domaine permettant l'oligomérisation de 53BP1 nécessaire pour une liaison complète de 53BP1 au niveau des dommages à l'ADN se situe dans sa partie N-terminal (Lottersberger et al., 2013).

2. 53BP1 et Rif1 orientent le choix de réparation des CDB

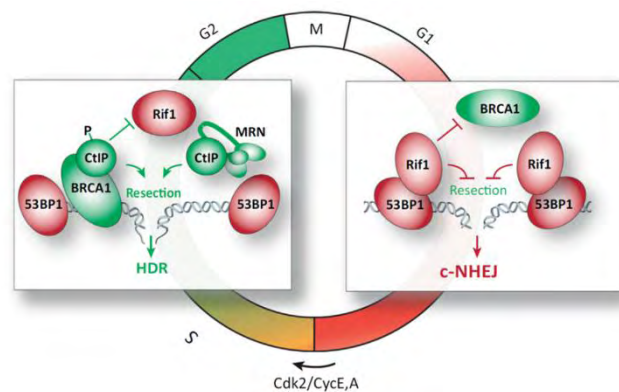


Figure 24 : Régulation dépendante du cycle cellulaire de la réponse aux CDB. Tiré de (Zimmermann and de Lange, 2013).

Le rôle de 53BP1 dans la réparation des CDB par c-NHEJ a été mis en évidence dans différents contextes, comme la diversification des IgG, les dysfonctionnements télomériques ou la déficience en BRCA1 (Pour revue (Zimmermann and de Lange, 2013)). Sans participer directement au c-NHEJ, 53BP1 favorise en fait cette voie de réparation en inhibant la résection des extrémités 5' des CDB et de ce fait la réparation par recombinaison homologe (Bunting et al., 2010; Lottersberger and Bothmer, 2013). Ce rôle de 53BP1 dans la promotion du NHEJ via l'inhibition de la résection est dépendante de Rif1, qui est essentiel à 53BP1 dans cette fonction (Zimmermann et al., 2013; Chapman et al., 2013; Escribano-Díaz et al., 2013; Feng et al., 2013). Le mode de recrutement de Rif 1 au CDB se fait via une interaction, certainement indirecte, avec les domaines de 53BP1 phosphorylés par ATM en réponse aux dommages à l'ADN, les sites ST/Q. En revanche, le mécanisme par lequel Rif1 pourrait contrôler la résection des CDB en 5' est inconnu. Enfin, cette régulation de la réponse aux CDB est régulée en fonction du cycle cellulaire (Figure 24). En phase G1, le NHEJ est donc privilégié au détriment de la RH par le couple 53BP1-Rif1 car Rif1 bloque le recrutement de BRCA1 à la chromatine, au niveau des cassures. En phase S/G2, à l'inverse, BRCA1 inhibe le

recrutement de Rif1 au CDB. Cette inhibition est dépendante de la phosphorylation de CtIP par les CDK, expliquant cette régulation de la réponse aux CDB en fonction du cycle cellulaire.

D. Problématique

Nous venons de voir qu'un stress réplicatif léger, endogène, ne générant pas l'activation de la réponse aux dommages à l'ADN pouvait induire des problèmes de réplication capables de traverser la phase S et d'entrer en mitose et d'être transmis aux cellules filles sous la forme de dommages à l'ADN assimilés à des CDB, protégés en phase G1 dans les foyers 53BP1. Cependant, le devenir de ces foyers, ou la façon dont sont pris en charge ces domaines ainsi que les conséquences sur le programme de réplication des cellules-filles sont inconnus. En particulier, nous nous sommes interrogés sur l'impact qu'ils pourraient avoir sur le *timing* de réplication. En effet, la mise en place du programme a lieu en début de phase G1, au moment du TDP ; et donc de façon concomitante avec la présence de ces foyers 53BP1, témoins des dommages transmis. La question est donc de savoir si ces régions cassées peuvent avoir une influence sur la régulation temporelle de l'activation des origines de réplication.

II. Résultats

1. Plan expérimental

Ce projet vise donc à évaluer les conséquences d'un stress réplicatif sur le *timing* de réplication au cours de la génération suivante. Pour répondre à cette question, nous avons mis au point le plan expérimental présenté Figure 25.

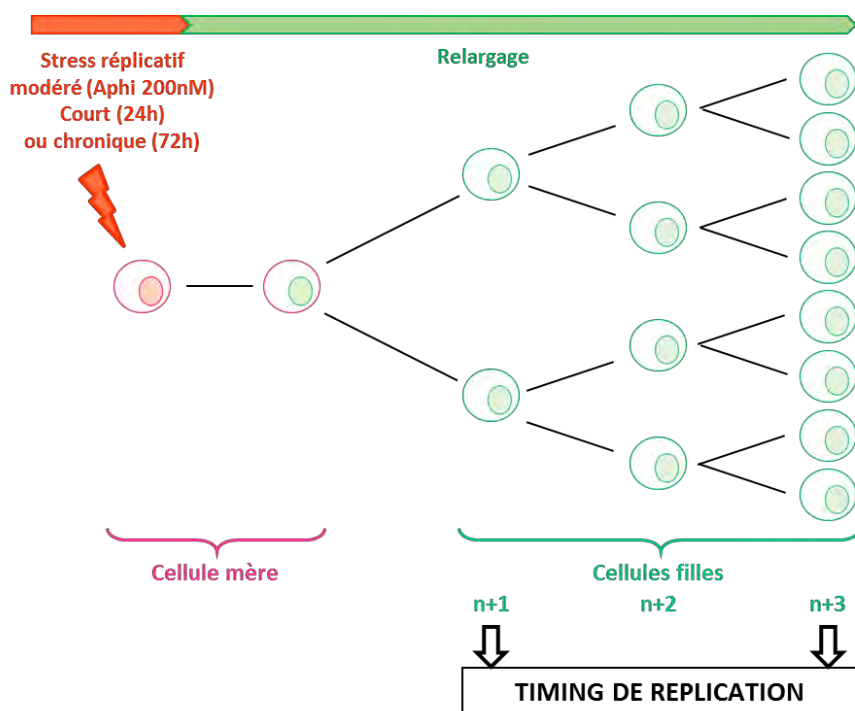


Figure 25 : Plan expérimental pour l'étude du *timing* de réplication dans les générations suivantes suite à un stress réplicatif. Les cellules ont été traitées par de l'Aphidicoline faible dose (200nM) pendant 24h ou 72h (cellules mère), puis remises en culture en absence de traitement. 15h ou 72h après, les cellules filles ont été récoltées pour réaliser l'étude du *timing* de réplication sur génome entier à la génération qui suit celle ayant subi le stress réplicatif (n+1) ou trois générations plus tard (n+3).

Les cellules ont d'abord été traitées avec une faible dose d'aphidicoline de 200 nM, qui comme observé dans de nombreuses études (Bergoglio et al., 2013; Koundrioukoff et al., 2013; Lukas et al., 2011), est trop faible pour induire l'activation du point de contrôle de phase S (Figure 26). En effet, que le traitement des cellules RKO soit court (24h) ou long (72h), le taux de Chk1 phosphorylé sur la sérine 345 (p-Chk1 (ser345)) est identique à celui détecté dans les cellules non traitées. En revanche, cette dose est suffisante pour induire la transmission d'ADN sous-répliqué aux cellules filles. En effet, après un traitement à l'aphidicoline 200nM pendant 24h, l'analyse par immunofluorescence des cellules interphasiques négatives pour la cycline A (Figure 27), et donc en G1, montre que le pourcentage de cellules présentant des foyers 53BP1 augmente consécutivement au traitement. La proportion de cellules en G1 présentant les foyers 53BP1 est encore plus importante lorsque le traitement a duré 72h (Figure 27C). En outre, le nombre de foyers 53BP1 par noyau dans les cellules négatives pour la cycline A augmente après le traitement, et ce d'autant plus que le traitement est long (Figure 27D). En revanche, lorsque les cellules ont été remises en culture en absence d'aphidicoline pendant 72h, la différence entre les conditions traitées et non traitées est perdue, que ce soit en terme de pourcentage de cellules

en phase G1 présentant des foyers 53BP1 qu'en terme de nombre de foyers par noyau (Figure 27C et D).

Le traitement à l'Aphidicoline à 200nM permet donc d'induire un stress réplcatif léger, à l'origine de dommages transmis aux cellules filles, sans activation du point de contrôle de phase S.

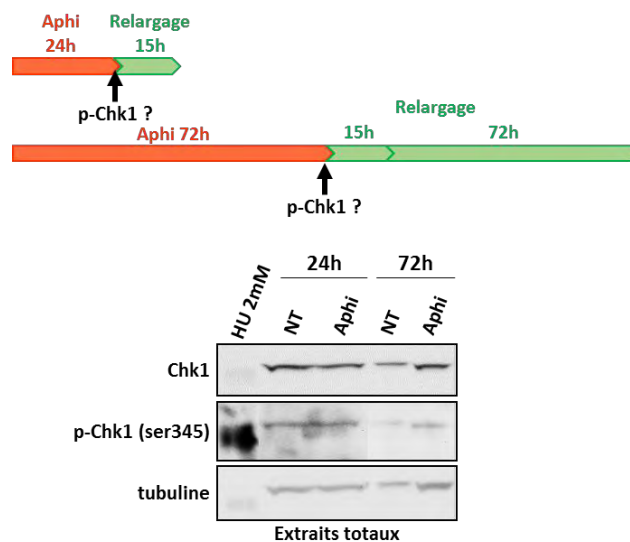


Figure 26 : Le traitement à l'aphidicoline 200nM n'induit pas d'activation de Chk1. Les cellules RKO ont été traitées à l'hydrouree (HU) 2mM pendant 1h, ou à l'aphidicoline (Aphi) 200nM pendant 24h ou 72h, ou non traitées (NT). L'analyse par western blot des extraits protéiques totaux a été réalisée avec les anticorps indiqués.

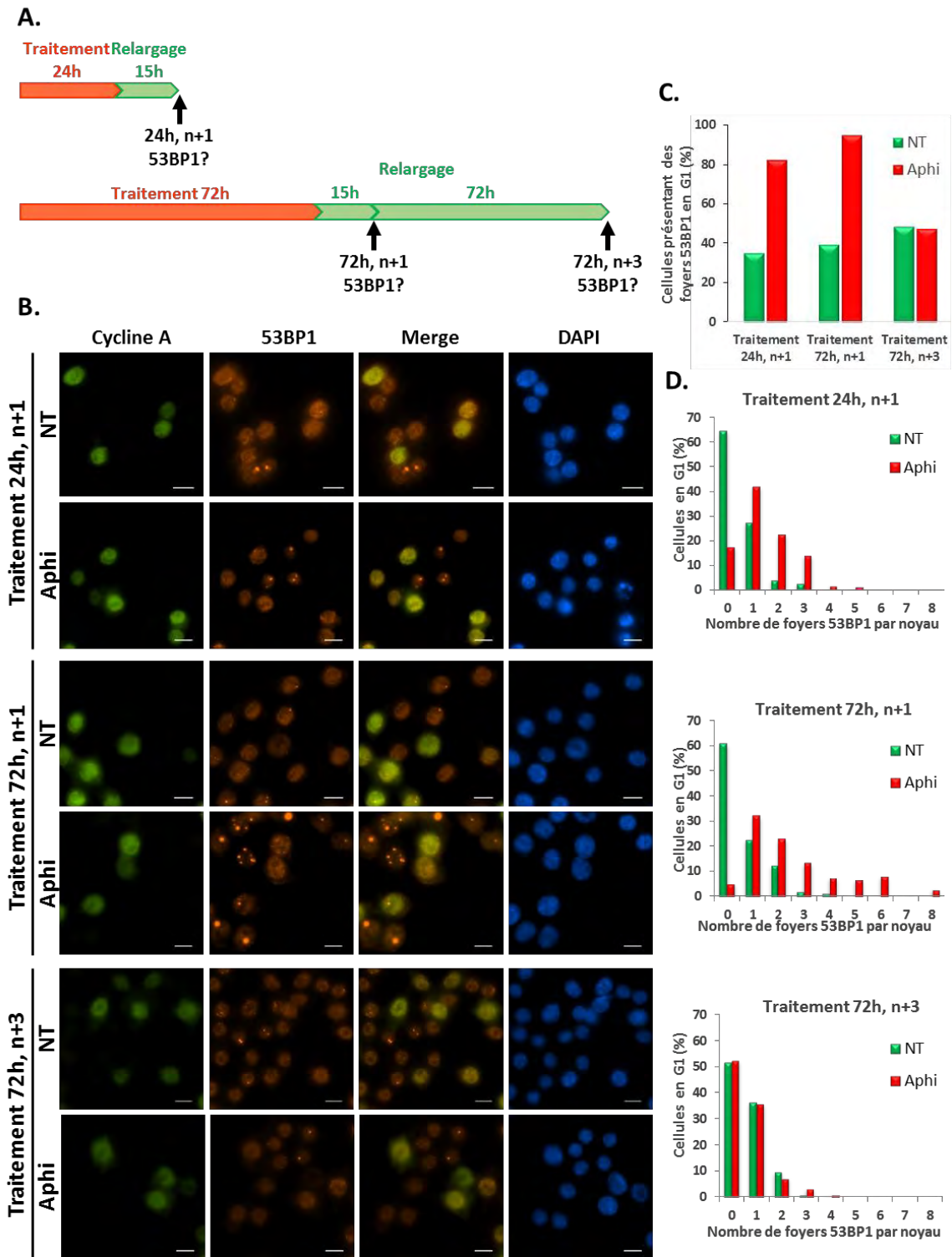


Figure 27 : Le stress répliatif léger induit l'augmentation des foyers 53BP1 en G1 dans les cellules filles. (A) Plan d'expérience : les cellules RKO ont été traitées à l'aphidicoline 200nM pendant 24h ou 72h puis relarguées dans le cycle cellulaire pour 15h (« Traitement 24h n+1 » et « Traitement 72h n+1 respectivement ») ou 72h (« Traitement 72h, n+3 »). (B) Images représentatives de foyers 53BP1 dans les cellules négatives pour la cycline A. (C) Histogrammes de quantification du pourcentage de cellules en G1 présentant des foyers 53BP1. (D) Histogrammes représentant la distribution du nombre de foyers 53BP1 par cellule en G. Aphi : Aphidicoline, NT : Non Traité

Les conditions de traitement établies, nous avons ensuite déterminé le temps de remise en culture sans stress des cellules mères, nécessaires pour l'étude du *timing* de réplication dans les cellules filles. Ce temps devait répondre à deux exigences. Il devait être suffisamment long pour que toutes les cellules mères ayant été en contact avec l'aphidicoline pendant la phase S ne soient plus en phase S au moment de la récolte des cellules dont le *timing* de réplication serait analysé. D'autre part, il devait être suffisamment court pour que toutes les cellules en phase S analysées soient bien issues des cellules mères ayant été stressées pendant toute ou partie de leur réplication. Pour déterminer la fenêtre pendant laquelle les cellules peuvent être récoltées, j'ai donc traité les cellules-mère à l'aphidicoline pendant 72h. A la fin de ces 72h, j'ai réalisé un pulse d'EdU, analogue de la thymidine, pour marquer les dernières cellules en phase S à avoir été traitées à l'aphidicoline. Après avoir ôté la drogue du milieu, j'ai réalisé une incorporation de BrdU à différents temps pour marquer les cellules en phase S au moment de la récolte. Une analyse par immunofluorescence a permis de visualiser les cellules ayant incorporé de l'EdU d'une part, c'est-à-dire les cellules mères traitées, et les cellules ayant incorporées du BrdU, donc en phase S au moment de la récolte d'autre part (Figure 28A et B). A t0h, près de 70% des cellules récoltées en phase S (BrdU positives) sont aussi EdU positives (Figure 28C). Ce pourcentage est un peu faible par rapport à celui attendu. En effet, à t0h, ce chiffre devrait être proche de 100%. Il se peut que des cellules en fin de phase S pendant le pulse EdU, et donc EdU positives, soient sorties de la phase S au moment du pulse BrdU. Elles seront alors EdU positives, mais BrdU négatives. Inversement, des cellules en fin de phase G1 au moment du pulse EdU, et donc EdU négatives peuvent être rentrée en phase S au cours du pulse BrdU, et seront donc EdU négatives mais BrdU positives. Comme je ne pense pas que ces deux dernières catégories puissent représenter 30% des cellules, je pense que le décalage vient d'une dénaturation incomplète de l'ADN avant de réaliser l'immunodétection anti-BrdU, induisant un signal un peu faible. Certaines cellules en début de phase S au moment de la récolte n'apparaissent donc peut-être pas positives au BrdU. Il serait donc plus rigoureux de réitérer l'expérience en co-incubant les cellules en présence d'EdU et de BrdU pour le temps t0h, et surtout en procédant à une dénaturation plus forte de l'ADN. Malgré ces réserves, cette cinétique nous a donc permis d'évaluer dans quelle phase du cycle sont les cellules de la génération suivant les cellules ayant subi un stress réplicatif léger, respectivement « Filles n+1 » et « Mère » (Figure 28D). Dès 9h, la quasi-totalité des cellules-mère ont quitté la phase S (moins de 20% de doubles positives) et sont donc soit en mitose, soit rentrées dans le cycle suivant.

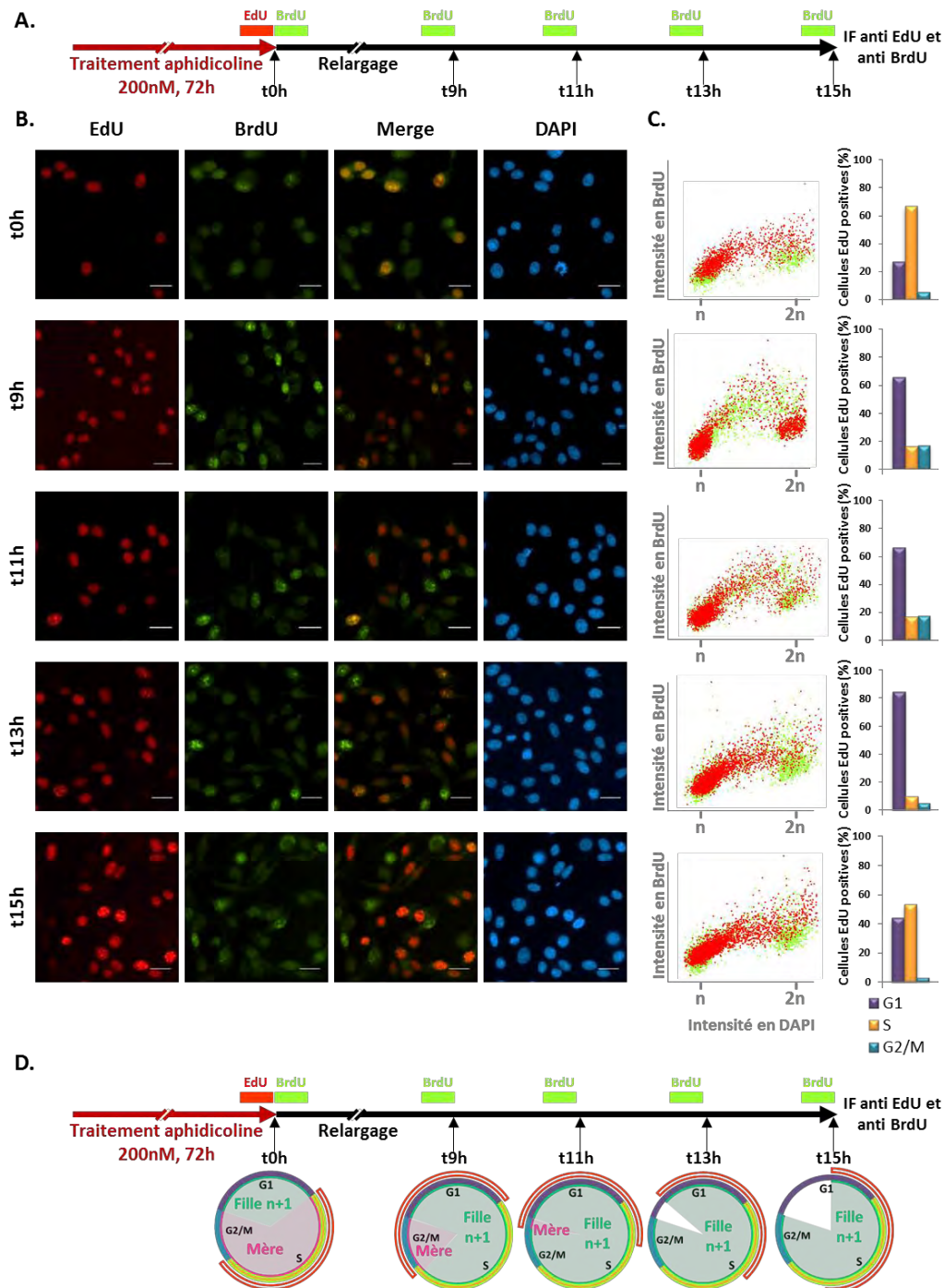


Figure 28 : Détermination du temps de relargage nécessaire pour la réalisation du *timing* de réplication à la génération $n+1$. (A) Plan expérimental. (B) Images représentatives de l'analyse par immunofluorescence des cellules marquées pour le BrdU et l'IdU aux différents temps de relargage du traitement aphidicoline. (C) Représentation graphique du cycle cellulaire des cellules récoltées aux différents temps (vert). Les cellules positives pour l'EdU sont marquées en rouge, et la répartition de cellules EdU positives dans le cycle des cellules récoltées à chaque temps est représentée par les histogrammes. (D) Représentation graphique de la position des cellules mères, traitées à l'aphidicoline, dans le cycle des cellules à chaque point de récolte. Le cercle représente le cycle cellulaire des cellules récoltées à chaque temps, les cellules en phase S, marquées au BrdU étant matérialisées par le double trait vert. Les cellules ayant été marquées à l'EdU sont matérialisées par le double trait rouge. Enfin, la position dans le cycle des cellules mères ayant été traitées à l'aphidicoline et les cellules filles de cellules mère traitées sont matérialisées par les diagrammes circulaires rose et vert respectivement. La part blanche correspond à des cellules filles de cellules non traitées. Echelle : $10\mu\text{M}$

A t11h, les cellules-mère ont presque toutes quitté les phases G2 et M et les cellules BrdU positives sont donc toutes des cellules filles n+1. Enfin, à 13 et 15h, les cellules-mères ont toutes passé la mitose et les cellules en phase S au moment de la récolte sont toutes des cellules filles n+1. Le *timing* de réplication peut donc être réalisé à ce moment-là. Pour mon plan expérimental, j'ai donc choisi de réaliser l'incorporation BrdU de 1h30, 13h30 après le relargage des cellules de l'aphidicoline, pour une récolte à t15h.

2. Le *timing* de réplication est avancé dans les cellules filles de cellules-mères ayant subi un stress réplicatif

a. *Modifications du timing de réplication*

L'analyse des profils de réplication sur l'ensemble du génome ont permis de mettre en évidence des modifications du *timing* de réplication dans les cellules-filles après que les cellules-mère aient été soumises à un stress réplicatif léger. Dans le cas d'un stress court comme d'un stress chronique, nous avons observé que 32 domaines tardifs dans la condition non traitée voient leur moment de réplication avancé à un moment intermédiaire (T→I) après stress réplicatif. Ils sont alors répliqués en milieu de phase S. De façon très intéressante, lorsque les cellules stressées ont été ensuite cultivées en absence d'aphidicoline sur plusieurs cycles (génération n+3), le profil de réplication se superpose avec celui de la condition contrôle, témoignant d'un retour à la normale du *timing* de réplication dans ces cellules (Figure 29A). De plus, lorsque le stress est appliqué sur plusieurs cycles (Traitement 72h), certaines régions précoces voient aussi leur *timing* de réplication avancé, alors que ce n'est pas le cas lorsque le stress n'a duré que 24h (Figure 29B). De même que précédemment, cette avancée de la réplication est perdue lorsque les cellules ont été remises en culture sans aphidicoline sur plusieurs cycles. Ces résultats montrent que suite à un stress réplicatif subi dans une cellule mère, le *timing* de réplication est modifié au cours de la phase S du cycle suivant. De plus cette modification semble d'autant plus importante que la durée du stress infligé est longue. En revanche, ces modifications ne sont plus observées aux générations suivantes pour lesquelles le *timing* de réplication redevient normal.

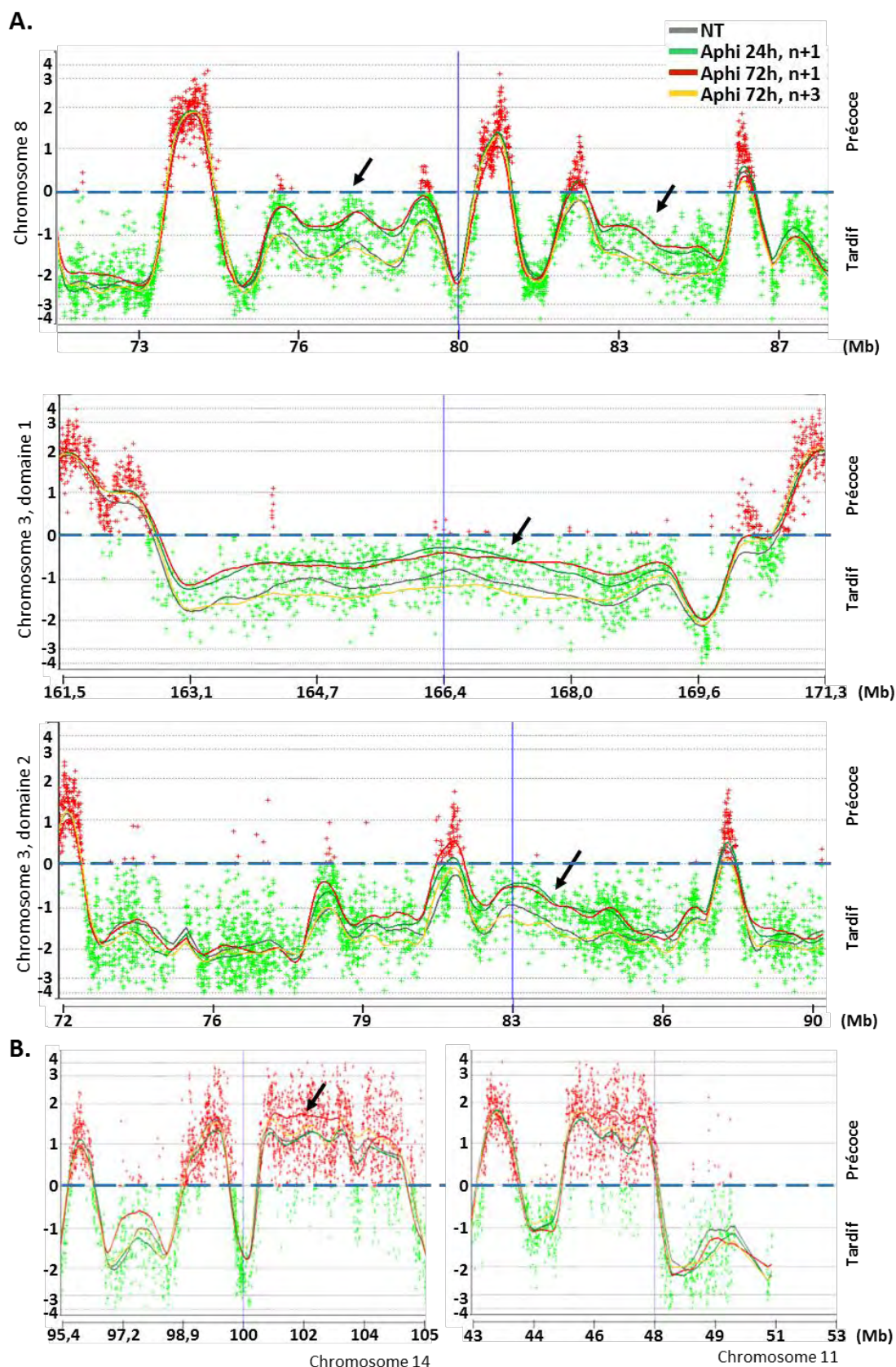


Figure 29 : Le *timing* de réplication est avancé dans les cellules-filles après un stress réplcatif appliqué aux cellules-mère. (A) Exemple de régions tardives dont le *timing* est avancé dans les cellules filles après un traitement court ou chronique des cellules mère à l'aphidicoline 200nM. (B) Exemple de régions précoces dont le *timing* de réplication est avancé après un temps long de traitement à l'aphidicoline.

b. Analyse des domaines affectés

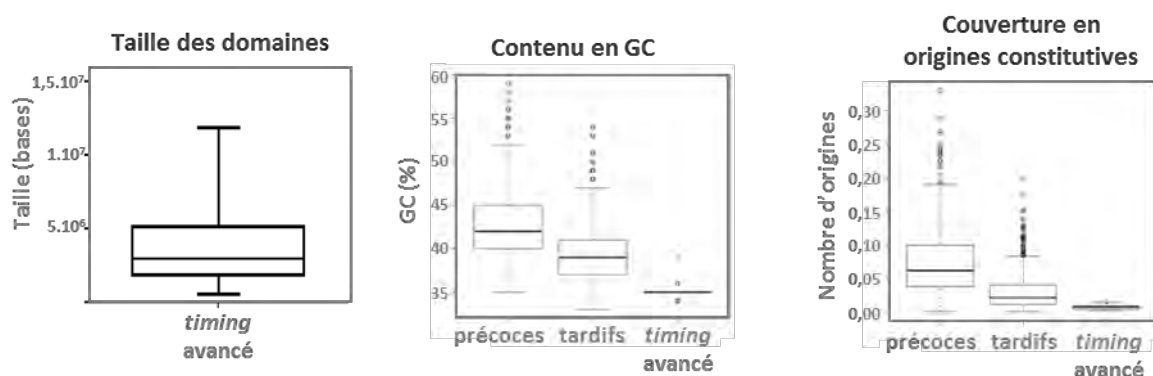


Figure 30 : Caractéristiques des régions tardives dont le *timing* de réplication est avancé dans les cellules-filles des cellules mères stressées.

Nous avons ensuite analysé les domaines tardifs dont le *timing* de réplication est avancé après les traitements de 24 et 48h des cellules-mère à l'aphidicoline dont des exemples sont présentés en Figure 29A. Ce sont de grands domaines, couvrant pour la plupart plusieurs mégabases (Mb), dont la taille médiane est de 2,9Mb (Figure 30, graphe de gauche). En accord avec la littérature (Costantini and Bernardi, 2008), dans notre expérience, les domaines tardifs présentent un contenu en GC plus faible que les domaines précoces. Ce pourcentage est encore plus bas dans les domaines T→I (Figure 30, graphe du milieu). Les origines de réplication étant associées à des séquences riches en GC, capables de former des structures G4, le faible contenu en GC de ces domaines dont le *timing* est modifié suggère qu'ils sont certainement déficitaires en origines de réplication. Pour tenter de répondre à cette question, nous nous sommes appuyés sur les travaux de Picard *et al*, qui ont identifié un certain nombre d'origines constitutives, communes à différentes lignées cellulaires (Picard et al., 2014). Et, pour chaque catégorie (précoce, tardif, domaines avancés), nous avons calculé la « couverture en origines constitutives », correspondant au rapport entre le nombre de bases couvertes par ces origines constitutives dans chaque catégorie, et le nombre total de bases couvertes par l'ensemble des domaines de la catégorie (Figure 30, graphe de droite). Nous avons observé que ces régions T→I ont une couverture en origines constitutives très faible, bien plus que les domaines précoces ou tardifs. On peut donc penser que ces domaines seront pauvres en origines, qu'elles soient constitutives ou non. Ces domaines dont le *timing* est avancé sont donc de grands domaines, riches en AT, et présentant une faible densité en origines de réplication.

Enfin, l'analyse des gènes contenus dans ces régions T→I révèle que la plupart d'entre eux participent à des processus biologiques impliqués dans les interactions de la cellule avec

son environnement, que ce soit en termes d'adhésion cellule à cellule, ou de cellule à la matrice extracellulaire (Annexe II).

c. Conclusion

L'ensemble de ces résultats suggère que lorsqu'une cellule a subi un stress réplicatif léger induisant des dommages pouvant être transmis aux cellules filles, le *timing* de réplication dans ces cellules-filles est modifié pour une sous population de domaines chromosomiques. Cette dérégulation temporelle de la réplication touche des régions particulières du génome. Elles sont caractérisées par leur grande taille, une représentation importante de nucléotides AT, donc probablement flexibles, et présentant une faible densité en origines. Ces caractéristiques, proches de celles des sites fragiles communs, suggèrent qu'elles doivent être particulièrement sensibles au traitement à l'aphidicoline et correspondrait donc à ces régions sous répliquées à l'origine des dommages transmis. Nous posons donc l'hypothèse que suite à un stress réplicatif subi dans une cellule-mère, des régions sous-répliquées sont transmises aux cellules filles sous une forme probablement cassée, et sont protégées dans les foyers 53BP1 en G1. Au cours de la phase S qui suit, il y a une avancée du *timing* de réplication au niveau de ces domaines, qui seront alors répliqués plus tôt pour donner le temps à cette cellule-fille de les répliquer complètement durant son cycle, et ceci alors qu'elle n'est elle-même pas soumise au stress réplicatif.

3. Mécanismes

a. Corrélation avec les foyers 53BP1

Si notre hypothèse est juste, cela signifie que les domaines qui sont protégées par les foyers 53BP1 en G1 sont ceux sur lesquels le *timing* de réplication est modifié. Cette hypothèse est suggérée par la corrélation que nous avons observée entre les changements de *timing* de réplication et les foyers 53BP1. En effet, suite au traitement à l'aphidicoline, nous avons observé une augmentation des foyers 53BP1 en G1 dans les cellules filles, que ce soit en termes de pourcentage de cellules en G1 présentant les foyers, ou de nombre de foyers par cellule en G1 (Figure 26). De plus, plus long était le traitement, plus l'induction de foyers 53BP1 était importante. De la même façon, nous avons observé des modifications du *timing* de réplication dans ces cellules-filles sur un plus grand nombre de domaines lorsque le traitement a été long (Figure 29). En revanche, au bout de trois cycles de relargage, la différence entre la condition non traitée et la condition traitée est perdue pour les foyers 53BP1 en G1 comme pour le *timing* de réplication (Figures 26 et 28).

Pour tester cette hypothèse, nous sommes en train de mettre au point au laboratoire deux approches expérimentales. Tout d'abord, nous voulons rechercher par hybridation *in situ* couplé à de l'immunofluorescence (immunoFISH) si les domaines T→I colocalisent avec les foyers 53BP1 en G1 après traitement aphidicoline à la génération précédente. De plus nous envisageons de réaliser des ChIP ciblées contre 53BP1 suivies d'une Q-PCR pour déterminer si nous immuno-précipitons les séquences dont le *timing* est modifié.

b. Recherche de l'implication de l'ADN polymérase Thêta

Comme nous l'avons vu dans le chapitre I des Résultats, l'ADN polymérase Thêta est un régulateur du *timing* de réplication. En outre, elle est recrutée à la chromatine tôt en G1 et participe à une voie de réparation des CDB fonctionnelle en G1, l'alt-NHEJ (Ceccaldi et al., 2015a; Fernandez-Vidal et al., 2014; Mateos-Gomez et al., 2015; Yousefzadeh et al., 2014). Ces arguments font d'elle un candidat potentiel dans cette modification du *timing* de réplication suite à un stress réplicatif survenu au cycle précédent. En effet, en intervenant dans la réparation des CDB transmises, elle peut influencer le *timing* de réplication du domaine chromosomique concerné à la phase S suivante. En ce sens, nous avons adopté la technique d'immunofluorescence de *Proximity Ligation Assay* qui permet de visualiser par immunofluorescence une grande proximité entre deux antigènes. En effet, elle repose sur l'utilisation d'anticorps secondaires couplés à des oligonucléotides qui peuvent s'hybrider entre eux s'ils sont suffisamment proches dans l'espace ($d < 40\text{nm}$). Une réaction d'amplification des sondes par RCA (*Rolling Circle Amplification*) en présence de dNTPs fluorescents est à l'origine d'un signal fluorescent visualisé sous forme de spot en microscopie à fluorescence. Utilisant cette technique, nous avons cherché une éventuelle interaction entre Polθ et 53BP1, mais aussi entre Polθ et Rif1, car, tout comme Polθ, cette protéine est un régulateur du *timing* de réplication et participe à une voie de réparation des CDB en G1, le c-NHEJ. Les premières expériences, réalisées dans des cellules asynchrones non traitées, ont montré une interaction entre Polθ et Rif1 (Figure 31A) et Polθ et 53BP1 (Figure 31B) qui semble spécifique, puisque le signal disparaissait après extinction de *POLQ* ou *RIF1* par ARN interférence. Nous cherchons actuellement à identifier ces interactions spécifiquement en phase G1 de cellules filles, et à évaluer si ces interactions sont intensifiées après un traitement à l'aphidicoline des cellules mères.

De plus des expériences de co-immunoprécipitation réalisées sur des extraits nucléaires de cellules en G1 montrent que Polθ coimmunoprécipite avec Rif1 et 53BP1, en particulier après traitement aphidicoline (Figure 31C).

cycle suivant, alors que le stress réplicatif n'est plus appliqué. Cette modification est réversible puisque trois générations après l'arrêt du stress, le *timing* de réplication revient à la normale. Ces modifications se traduisent par une avancée du *timing* de réplication sur certains domaines. Elles concernent majoritairement des domaines particuliers du génome, caractérisés par leur réplication tardive, leur richesse en AT et leur possible pauvreté en origines. Notre hypothèse est donc que ces régions, qui présentent des caractéristiques communes avec les SFC, sont préférentiellement touchées par le stress réplicatif appliqué dans les cellules-mères, induisant des dommages transmis aux cellules-filles comme en témoigne la présence de gros foyers 53BP1 en phase G1. Ces cellules-filles vont alors avancer leur programme temporel d'activation des origines de réplication dans ces domaines pour optimiser les chances d'une réplication complète au cours de leur phase S. Cette adaptation pourrait éventuellement être médiée par Polθ ou Rif1, deux régulateurs décrits du *timing* de réplication, qui participent à des voies de réparation des CDB actives en phase G1, le MMEJ et le c-NHEJ respectivement. En effet, la forme des lésions transmises en G1 n'est pas précisément définie. Le recrutement de 53BP1 et de γH2AX suggère que ce sont des CDB, mais on ne connaît pas la nature de leurs extrémités. Compte-tenu de la façon dont elles sont générées, il est fort possible qu'il y ait d'ailleurs une grande hétérogénéité et complexité dans ces types de dommages transmis. D'un côté, le couple 53BP1-Rif1 inhibe la résection en G1 et le c-NHEJ a comme substrat des CDB à extrémités essentiellement franche. D'un autre côté, une étude a montré que 53BP1 peut favoriser le MMEJ en G1 (Xiong et al., 2015). On peut donc émettre l'hypothèse que 53BP1 puisse aussi permettre le recrutement de Polθ dans le cadre d'une réparation par MMEJ. L'intervention dans la réparation des dommages de Rif1 ou Polθ, les deux régulateurs du *timing* de réplication connus à ce jour, pourrait alors impacter le programme de réplication de la phase S suivante (Figure 32).

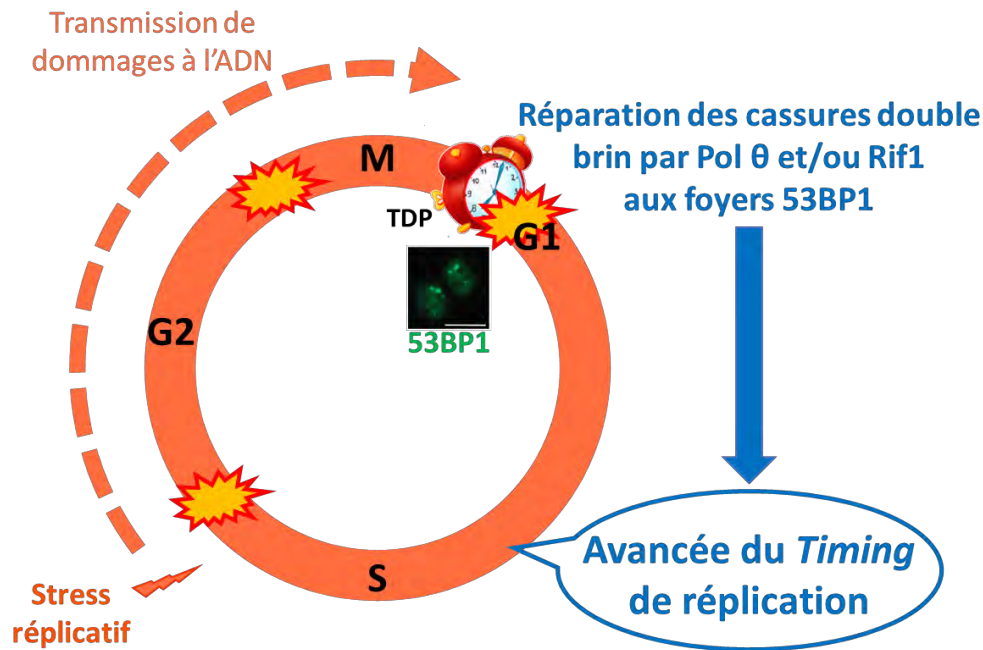


Figure 32 : Modèle pour l'adaptation du *timing* de réplication induite par un stress réplicatif du cycle précédent. TDP= *Timing Decision Point*

B. Perspectives

Cette observation selon laquelle le *timing* de réplication est modifié suite à un stress réplicatif du cycle précédent pose de nombreuses questions.

Tout d'abord, cette modification est-elle vraiment une adaptation aux dommages transmis suite à ce stress ? Comme évoqué précédemment, une des premières choses à faire est d'évaluer si les changements de *timing* de réplication ont lieu dans des régions recouvertes par 53BP1 en G1, et donc au niveau de lésions transmises induites par le stress réplicatif. Pour cela, de l'immuno-FISH ainsi que le ChIP nous permettront peut-être d'évaluer la présence de 53BP1 sur ces régions en G1. De plus, pour évaluer si la présence de 53BP1 joue un rôle dans ces modifications du *timing* de réplication, nous envisageons de réaliser des expériences similaires en inhibant le recrutement aux dommages de 53BP1 en G1. Nous avons vu que la phosphatase PP4C ainsi que sa sous-unité régulatrice PP4CR3 β sont responsables de la déphosphorylation de 53BP1 à sa sortie de mitose, permettant ainsi son recrutement au niveau des dommages transmis (Lee et al., 2014). Nous envisageons donc d'éteindre par siRNA le complexe à activité phosphatase PP4C/3R β , soit par extinction simultanée des deux protéines, soit par l'extinction d'une seule des deux. En particulier, cibler uniquement PP4C3R β peut permettre de s'affranchir des larges conséquences que pourraient avoir l'extinction de PP4C, phosphatase impliquée dans d'autres mécanismes de réponse aux dommages à l'ADN. Si nous perdons les modifications du *timing* de réplication précédemment observées, cela

montrera que ces modifications sont dépendantes de 53BP1. Par ailleurs, les rôles éventuels de Polθ et Rif1 doivent être explorés. Tout d'abord, les expériences de PLA et d'IP doivent être poursuivies. Le PLA en particulier doit être réalisé dans des cellules en G1, après traitement aphidicoline, de façon à voir si les interactions entre 53BP1 et Polθ ou Rif1 sont renforcées après le traitement. Si c'est le cas, nous proposerions que 53BP1 puisse servir de plateforme au recrutement des protéines de la réparation de ces lésions transmises en G1. La combinaison d'un PLA ciblant Polθ et Rif1 couplée à une immunofluorescence anti-53BP1 et Cycline A permettrait aussi de déterminer si l'interaction entre Polθ et Rif1 a lieu au sein de ces foyers 53BP1 en G1. Enfin, l'implication de Polθ dans la modification du *timing* de réplication via son activité de réparation peut-être mise en évidence par l'utilisation de mutants dans le domaine polymérase, dont on sait qu'il est essentiel à son rôle dans le MMEJ.

Par ailleurs, nous avons vu que le *timing* de réplication est fortement influencé par l'environnement chromatinien. On peut donc aussi émettre l'hypothèse que cette modification du *timing* de réplication suite à un stress répliatif du cycle précédent soit liée à des modifications d'histones particulières au niveau de ces domaines. Il serait donc intéressant d'explorer les marques d'histones qui sont associées aux domaines dont le programme temporel est modifié. En effet, on peut envisager que ces domaines, subissant des problèmes de réplication ne présentent pas les mêmes modifications d'histones que la chromatine correctement dupliquée. Tout d'abord, au cours de la réplication, les histones sont momentanément dissociées de la chromatine par les hélicases répliatives. Une fois l'ADN dupliqué, les nucléosomes se réorganisent à partir des histones parentales et de nouvelles histones dites naïves, portant chacune des modifications différentes. Au cours de la maturation de la chromatine, les histones naïves vont perdre leurs marques spécifiques pour adopter celles des histones parentales, permettant la transmission de l'information épigénétique (Pour revue (Margueron and Reinberg, 2010)). En cas de stress répliatif, ce renouvellement des histones peut être perturbé, induisant une modification de cette information épigénétique transmise (Cea et al., 2015; Jasencakova et al., 2010; Khurana and Oberdoerffer, 2015). On peut donc imaginer que le stress répliatif appliqué dans les cellules-mères induise des modifications des marques d'histones sur les domaines problématiques, expliquant un changement de *timing* de réplication au cycle suivant. Par ailleurs, comme nous l'avons vu, le recrutement de 53BP1 aux CDB est aussi régulé par des modifications post-traductionnelles des histones, dont entre-autres H4K20me2 ou la déacétylation de H4K16, marques décrites dans la régulation du *timing* de réplication (Cf Introduction-Chapitre II-Partie II). La prise en charge des lésions transmises dans les foyers 53BP1 dès l'entrée en

phase G1, et donc de façon concomitante avec le TDP, est certainement associée à des modifications d'histones qui pourraient influencer le *timing* de réplication à la phase S suivante.

Enfin, un aspect intéressant à développer est d'explorer les conséquences au niveau cellulaire de ces modifications de *timing* de réplication. Tout d'abord, nous avons vu que les gènes contenus dans les régions affectées appartiennent à des processus biologiques en rapport avec l'adhérence cellulaire. On peut donc se demander si leur expression est affectée, et si les propriétés d'adhésion ou de migration des cellules filles peuvent être affectées. De plus, nous avons vu que les domaines avancés présentent des caractéristiques proches des sites fragiles. Il serait intéressant de voir si ce sont effectivement des sites fragiles en évaluant leur expression par des expériences de FISH sur étalement métaphasique. De plus, si nous trouvons des conditions empêchant la modification du *timing* de réplication nous pourrions évaluer si la modification de *timing* de réplication permet de préserver la stabilité de ces sites. Ceci serait très intéressant dans un contexte tumoral dans lequel le stress réplcatif est chronique et où cette modification du programme temporel des origines de réplication pourrait constituer un mécanisme d'adaptation des cellules cancéreuses à leur propre stress, ou au stress induits par les agents anti-tumoraux (chimiothérapies ciblant la réplication ou radiothérapies).

CONCLUSIONS ET **DISCUSSION GÉNÉRALE**

I. Polθ régule le *timing* de réplication

Les travaux réalisés au cours de ma thèse ont tout d'abord permis d'identifier, après Rif1, un deuxième régulateur du *timing* de réplication dans les cellules humaines, l'ADN polymérase Thêta. Polθ est recrutée à la chromatine en tout début de phase G1, de façon concomitante avec le TDP. La co-immunoprécipitation avec ORC2 et ORC4 suggère qu'elle se situe au niveau des origines de réplication. Enfin, son absence induit une augmentation dans le recrutement des Mcm la chromatine en G1, suggérant qu'elle influence la mise en place du pré-RC. L'ensemble de ces résultats pose de nombreuses questions. Tout d'abord, on peut se demander quelles sont les origines concernées et quels sont les mécanismes moléculaires mis en jeu. Le fait qu'une ADN polymérase spécialisée puisse réguler le *timing* de réplication était, à ma connaissance, totalement inédit et reste assez surprenant. En effet, la nécessité d'une protéine capable d'activité polymérase dans la régulation temporelle de l'activation des origines de réplication est contre-intuitive, étant donné que ce programme est établi dans une phase du cycle où la synthèse d'ADN est restreinte. Comme évoqué précédemment, on peut imaginer que cette activité est néanmoins nécessaire en phase G1, dans des processus de réparation des dommages à l'ADN. Il se peut aussi que Polθ participe, via son activité polymérase à l'activation d'une sous-catégorie d'origines, dont elle régule le *timing* de réplication, au cours de la phase S. Par exemple, la combinaison de ses activités polymérase et hélicase lui confèreraient les capacités d'ouvrir la double hélice d'une part, et synthétiser de l'ADN d'autre part. Ce rôle pourrait en particulier être pertinent dans des régions du génome dont la séquence favorise la formation d'ADN structuré, et où la réunion de l'hélicase et de la polymérase en une seule entité fonctionnelle permettrait une efficacité de démarrage à partir d'un site d'initiation. Pour le moment, le choix des origines ciblées par Polθ reste un mystère. Les expériences de ChIPseq pourront peut-être donner quelques réponses à cette question et permettre d'identifier une ou plusieurs spécificités de séquences de liaison de Polθ à la chromatine. En particulier, il sera intéressant de voir si elle est recrutée au niveau de régions capables de former des G-quadruplex, structures associées aux origines de réplication. En effet, il a été proposé que la présence de ces G4 aux origines joue un rôle topologique, en influençant la structure de la chromatine pour imposer une conformation ouverte qui favoriserait le recrutement des protéines du pré-RC à la chromatine. Polθ, en déstabilisant ces G4, limiterait le recrutement des Mcm au niveau des origines, régulant ainsi le *timing* de réplication. Pour répondre à cette question, une approche expérimentale consisterait à comparer les ChIP anti-Polθ avec des ChIP anti-MCM en présence ou en

absence de Polθ sur des cellules synchronisées en G1. En effet, si l'hypothèse d'une compétition entre Polθ et les MCM au niveau de certaines origines est vraie, nous devrions observer, pour les pics d'enrichissement dans les ChIP Polθ un pic d'enrichissement dans les ChIP Mcm en absence de Polθ par rapport à la condition contrôle.

Par ailleurs, nous avons vu qu'un volet important de la régulation du *timing* de réplication correspond à l'organisation de la chromatine dans le noyau. Or l'absence de Polθ induit un raccourcissement des boucles de chromatine, suggérant que Polθ influence cette organisation nucléaire. Dans ce contexte, il serait d'une part intéressant d'évaluer si les régions dont le *timing* de réplication est modifié sont relocalisées dans le noyau en absence de Polθ. Et pour comprendre comment Polθ pourrait influencer l'architecture nucléaire, il faudrait déterminer où elle se situe dans le noyau, en particulier si elle se trouve au niveau de la matrice nucléaire. On peut pour cela réaliser des expériences de fractionnement subcellulaire, dans lequel une digestion de l'ADN à la DNase permettra de distinguer les fractions chromatiniennes solubles de celles attachées à la matrice nucléaire. Des expériences d'immunofluorescence peuvent aussi nous permettre de localiser Polθ, ou bien de voir si elle colocalise avec les lamines, situées à la matrice nucléaire. Le fait que Polθ co-immunoprécipite avec Rif1, qui lui-même colocalise avec la lamine B1 et se situe dans la fraction nucléaire insoluble (Yamazaki et al., 2012), suggère qu'au moins une partie de Polθ chromatinien est probablement lui aussi ancré à la matrice nucléaire, régulant ainsi la taille des boucles de chromatine.

Un autre point qui me paraît extrêmement intéressant à explorer est le lien qu'il peut y avoir entre Polθ et les processus de différenciation ou de développement. En effet, le *timing* de réplication est directement lié à l'identité de la cellule. Il évolue avec le stade de différenciation, pendant lequel les domaines courts dans les cellules peu différenciées sont consolidés en domaines de réplication plus grands (Hiratani et al., 2008; Ryba et al., 2011b). Or, dans nos expériences, les profils de *timing* de réplication obtenus après déplétion de Polθ sont apparentés à ceux d'une cellule différenciée. En outre, l'absence de Tebichi, l'homologue de Polθ chez la plante *A. Thaliana* entraîne des problèmes de différenciation dans le méristème (Inagaki et al., 2006). On peut donc se demander si Polθ ne pourrait pas jouer un rôle dans la différenciation dans un modèle humain et il serait intéressant d'étudier son expression dans des systèmes de différenciation *in vitro*, ainsi que de voir quel impact pourrait avoir l'extinction de Polθ sur ce processus. L'étude de l'expression des gènes dans les régions dont le *timing* de réplication est affecté, en fonction du statut de Polθ pourrait aussi apporter des réponses à ce sujet.

II. Modification du *timing* de réplication après un stress réplicatif à la génération précédente.

La deuxième partie de mes travaux de thèse révèle un processus biologique tout à fait inédit : une modification du *timing* de réplication en absence de stress réplicatif, mais en réponse à un stress subi au cycle précédent. Une hypothèse très séduisante est que les cellules-filles ont en quelques sortes « gardé en mémoire » le stress réplicatif subi par leur mère, et adapté leur *timing* de réplication préventivement au cours de leur phase S pour anticiper les problèmes de réplication sur certaines régions. Or, il semble qu'en phase G1, 53BP1 puisse être en complexe avec deux régulateurs du *timing* de réplication, qui sont également acteurs de la réparation des CDB, Polθ et Rif1. Ce résultat suggère que cette modification de l'activation des origines de réplication puisse être médié par le couplage entre la réparation de ces dommages en G1 et le TDP, qui a lieu au moment où Polθ et Rif1 sont recrutés à la chromatine. On ne peut cependant pas exclure que ce couplage entre réparation des dommages transmis et activation des origines ait lieu pendant la phase S, dans lequel la réparation aurait lieu plus tôt dans la phase S que le moment d'activation des origines de réplication normal de ces domaines induisant leur réplication plus précoce.

Si l'avancée du *timing* de réplication observée correspond effectivement à une adaptation des cellules-filles par l'avancée de la réplication des régions sensibles au stress réplicatif des cellules-mère, cela pourrait avoir des significations biologiques importantes. Tout d'abord, dans un contexte tumoral, plusieurs études ont montré que le stress réplicatif est chronique, élevé, et participe au processus de carcinogénèse en amplifiant l'instabilité génétique. Or, nous avons aussi vu que la cellule dispose normalement de voies de réponses à ce stress réplicatif qui peuvent conduire jusqu'à la mort cellulaire si le stress réplicatif n'est pas contenu. Pourtant, les cellules tumorales survivent à ce stress réplicatif intrinsèque. Ce paradoxe suggère que ces cellules ont développé des moyens de résistance et d'adaptation qui leur permettent de survivre dans ces conditions. L'adaptation du *timing* de réplication en réponse à un stress subi dans les générations précédentes pourrait constituer un de ces moyens d'adaptation des cellules tumorales au stress réplicatif au cours d'une expansion clonale. Ainsi, l'avancée de la réplication des domaines sensibles au stress réplicatif permettrait de limiter, mais probablement pas empêcher, l'instabilité génétique qui pourrait être générée à leur niveau. Plus le stress est long (24h *versus* 72h dans nos expériences), plus il touche un nombre important de domaines et les modifications du *timing* de réplication sont d'autant plus importantes. Cette hypothèse pourrait expliquer les avancées du *timing* de réplication

observées dans les cellules leucémiques lymphoblastiques, en comparaison avec des cellules normales (Ryba et al., 2012). De plus, l'existence d'une telle réponse au stress réplicatif pourrait aussi avoir des conséquences en termes de réponse aux agents chimio-thérapeutiques qui ciblent la réplication. En effet, ce processus pourrait aussi constituer un moyen de résistance à ces drogues, en limitant les effets sur l'instabilité génétique, et donc en allant à l'encontre de l'induction de la mort cellulaire escomptée. Il serait ainsi intéressant de réaliser des expériences de *timing* de réplication dans des conditions mimant les protocoles d'administration des chimiothérapies. Ces protocoles reposent en général sur l'administration du médicament en plusieurs fois, avec des temps de repos entre chaque administration. En suivant un protocole similaire pour le traitement de cellules cancéreuses en culture, nous pourrions tout d'abord voir si l'adaptation du *timing* de réplication observée après traitement aphidicoline est conservée avec une drogue différente, et quelles sont les caractéristiques des régions affectées. Sur le long terme, nous pourrions évaluer s'il y a une modification globale du *timing* de réplication en réponse à ce stress, ou si à chaque fois, le *timing* de réplication revient à sa situation initiale, sans conservation d'aucune variation au fur et à mesure de la répétition du traitement.

III. Polθ, acteur dans l'adaptation du *timing* de réplication au stress réplicatif ?

L'implication éventuelle de Polθ dans ce processus d'adaptation des cellules au stress réplicatif est très intéressante au regard de sa signature particulière dans de nombreux cancers, sa surexpression étant facteur de mauvais pronostic dans plusieurs tumeurs solides (Figure 33A). En effet, si l'hypothèse selon laquelle cette ADN polymérase participe à cette adaptation du programme temporel d'activation des origines en cas de stress est vraie, alors la surexpression de Polθ dans ces cancers peut tout d'abord constituer un des moyens de résistance des cellules tumorales à leur propre stress réplicatif. En effet, au cours du développement tumoral, le stress réplicatif intrinsèque de la cellule cancéreuse est transmis de génération en génération au cours de l'expansion clonale des cellules tumorales, d'autant que les checkpoint y sont le plus souvent défectueux. Dans ce cadre, la surexpression de Polθ permettrait de renforcer la réponse et l'adaptation des cellules au stress réplicatif, non seulement *via* la modulation du *timing* de réplication, mais aussi de par le rôle de Polθ dans la réparation des dommages à l'ADN dans la voie du alt-NHEJ. En ce sens, Polθ participerait au développement de la maladie. Dans un second temps, cette surexpression peut aussi permettre aux cellules de mieux résister aux traitements contre le cancer qui ciblent la réplication (chimiothérapies et/ou radiothérapies). Au sein d'une tumeur hétérogène, les cellules

surexprimant Polθ seront alors sélectionnées pendant le traitement et peuvent être à l'origine de rechute ou d'une plus grande agressivité des tumeurs (Figure 33B).

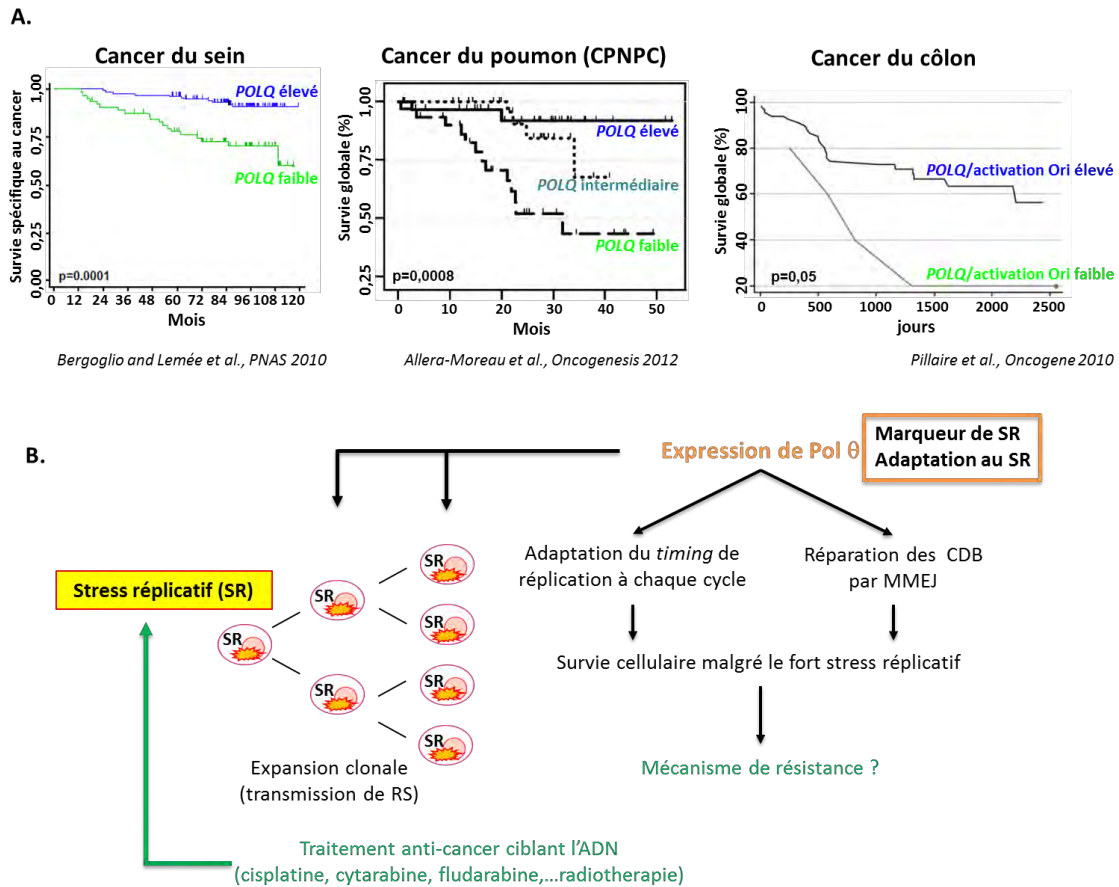


Figure 33 : La surexpression de Polθ dans les cancers, capacité de survie en conditions de stress réplicatif chronique ? (A) La surexpression de Polθ est un facteur de mauvais pronostic dans les cancers du sein, du poumon non à petites cellules (CPNPC) et du colon lorsque cette surexpression est associée à celles de gènes impliqués dans l'activation des origines de réplication (activation Ori). **(B)** Modèle de l'adaptation de la cellule tumorale à son propre stress réplicatif (SR) chronique, ou à celui induit par les traitements anti-cancéreux ciblant la réplication, par la surexpression de Polθ.

Ainsi, le ciblage de Polθ pourrait constituer une stratégie thérapeutique importante, non seulement dans le cadre d'une stratégie de létalité synthétique dans le cas de tumeurs déficientes pour la RH (Ceccaldi et al., 2015a; Mateos-Gomez et al., 2015), mais aussi dans les autres tumeurs, pour limiter les capacités d'adaptation des cellules au stress réplicatif intrinsèque ou induit par les traitements. Cependant, l'observation selon laquelle la déplétion de cette ADN polymérase dans des cellules B chez la souris induit une augmentation des translocations au niveau du locus des immunoglobulines *IgH* (Yousefzadeh et al., 2014) et de l'instabilité génétique chez *C. elegans* (Koole et al., 2014; Roerink et al., 2014) suggèrent que le ciblage de Polθ dans le traitement du cancer doit être prudemment envisagée.

ANNEXES

I. ANNEXE 1 : Classification des gènes situés dans des régions dont le *timing* de réplication est modifié après déplétion de *POLO* dans les cellules RKO.

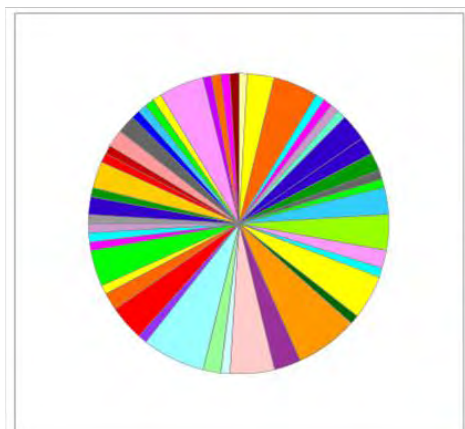
1. En fonction des voies de signalisation

Régions précoces -> tardives



5-Hydroxytryptamine biosynthesis (P04371)	
SHT1 type receptor mediated signaling pathway (P04373)	
Alzheimer disease-amyloid secretase pathway (P00003)	
Alzheimer disease-presenilin pathway (P00004)	
Aminobutyrate degradation (P02726)	
Angiogenesis (P00005)	
Apoptosis signaling pathway (P00006)	
Asparagine and aspartate biosynthesis (P02730)	
Axon guidance mediated by netrin (P00009)	
Axon guidance mediated by semaphorins (P00007)	
B cell activation (P00010)	
Beta1 adrenergic receptor signaling pathway (P04377)	
Beta2 adrenergic receptor signaling pathway (P04378)	
Blood coagulation (P00011)	
Cadherin signaling pathway (P00012)	
Cytoskeletal regulation by Rho GTPase (P00016)	
Dopamine receptor mediated signaling pathway (P05912)	
EGF receptor signaling pathway (P00018)	
Endothelin signaling pathway (P00019)	
Enkephalin release (P05913)	
FAS signaling pathway (P00020)	
FGF signaling pathway (P00021)	
GABA-B receptor II signaling (P05731)	
Gamma-aminobutyric acid synthesis (P04384)	
Gonadotropin releasing hormone receptor pathway (P06664)	
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)	
Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028)	
Histamine H2 receptor mediated signaling pathway (P04386)	
Huntington disease (P00029)	
Hypoxia response via HIF activation (P00030)	
Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	
Insulin/IGF pathway-protein kinase B signaling cascade (P00033)	
Integrin signalling pathway (P00034)	
Interleukin signaling pathway (P00036)	
Ionotropic glutamate receptor pathway (P00037)	
Metabotropic glutamate receptor group I pathway (P00041)	
Metabotropic glutamate receptor group II pathway (P00040)	
Metabotropic glutamate receptor group III pathway (P00039)	
Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042)	
Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043)	
Nicotine pharmacodynamics pathway (P06587)	
Nicotinic acetylcholine receptor signaling pathway (P00044)	
Notch signaling pathway (P00045)	
PDGF signaling pathway (P00047)	
PI3 kinase pathway (P00048)	
Ras Pathway (P04393)	
Serine glycine biosynthesis (P02776)	
T cell activation (P00053)	
TGF-beta signaling pathway (P00052)	
Thyrotropin-releasing hormone receptor signaling pathway (P04394)	
Toll receptor signalling pathway (P00054)	
Transcription regulation by bZIP transcription factor (P00055)	
VEGF signaling pathway (P00056)	
Wnt signaling pathway (P00057)	
p53 pathway feedback loops 2 (P04398)	
p53 pathway (P00059)	

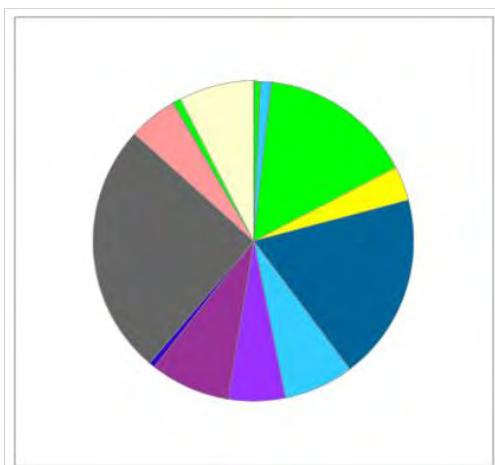
Régions tardives -> précoces



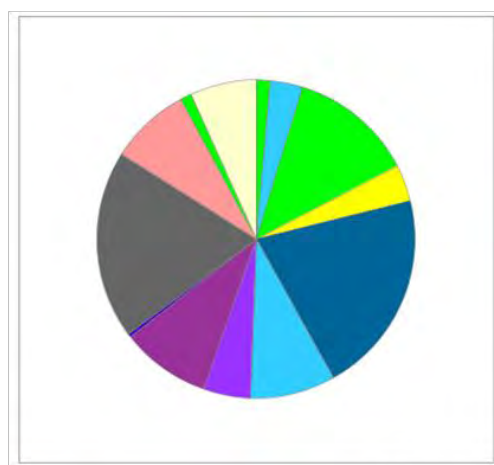
5HT4 type receptor mediated signaling pathway (P04376)	
Alzheimer disease-presenilin pathway (P00004)	
Angiogenesis (P00005)	
Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911)	
Apoptosis signaling pathway (P00006)	
Axon guidance mediated by Slit/Robo (P00008)	
Beta2 adrenergic receptor signaling pathway (P04378)	
Blood coagulation (P00011)	
CKKR signaling map (P06959)	
Cadherin signaling pathway (P00012)	
Cytoskeletal regulation by Rho GTPase (P00016)	
De novo pyrimidine deoxyribonucleotide biosynthesis (P02739)	
EGF receptor signaling pathway (P00018)	
Endothelin signalling pathway (P00019)	
FGF signaling pathway (P00021)	
GABA-B receptor II signaling (P05731)	
Gonadotropin releasing hormone receptor pathway (P06664)	
Hedgehog signaling pathway (P00025)	
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)	
Huntington disease (P00029)	
Hypoxia response via HIF activation (P00030)	
Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	
Integrin signalling pathway (P00034)	
Interleukin signaling pathway (P00036)	
Ionotropic glutamate receptor pathway (P00037)	
Metabotropic glutamate receptor group I pathway (P00041)	
Metabotropic glutamate receptor group II pathway (P00040)	
Metabotropic glutamate receptor group III pathway (P00039)	
Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042)	
Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043)	
Nicotinic acetylcholine receptor signaling pathway (P00044)	
Notch signaling pathway (P00045)	
PDGF signaling pathway (P00047)	
PI3 kinase pathway (P00048)	
Parkinson disease (P00049)	
Ras Pathway (P04393)	
Synaptic vesicle trafficking (P05734)	
T cell activation (P00053)	
TGF-beta signaling pathway (P00052)	
Thyrotropin-releasing hormone receptor signaling pathway (P04394)	
Toll receptor signaling pathway (P00054)	
Vasopressin synthesis (P04395)	
Vitamin D metabolism and pathway (P04396)	
Wnt signaling pathway (P00057)	
p38 MAPK pathway (P05918)	
p53 pathway by glucose deprivation (P04397)	
p53 pathway feedback loops 2 (P04398)	
p53 pathway (P00059)	

2. En fonction des processus biologiques

Régions précoces -> tardives

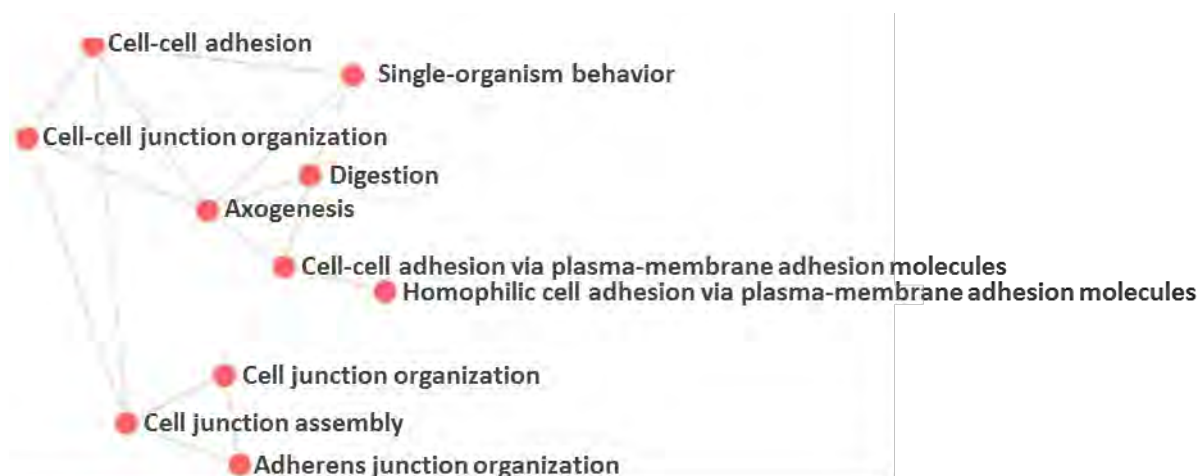


Régions tardives -> précoces



	apoptotic process (GO:0006915)	
	biological adhesion (GO:0022610)	
	biological regulation (GO:0065007)	
	cellular component organization or biogenesis (GO:0071840)	
	cellular process (GO:0009987)	
	developmental process (GO:0032502)	
	immune system process (GO:0002376)	
	localization (GO:0051179)	
	locomotion (GO:0040011)	
	metabolic process (GO:0008152)	
	multicellular organismal process (GO:0032501)	
	reproduction (GO:0000003)	
	response to stimulus (GO:0050896)	

II. ANNEXE II : Classification des gènes contenus dans les régions tardives dont le timing de réplication est avancé dans les cellules-filles des cellules mères stressées



PROTOCOLES

Expériences d'ImmunoFluorescence

Conditions de culture

Les cellules RKO ont été cultivées dans du DMEM complété par 10% de Sérum de Veau Fœtal (SVF). Les conditions standard d'incubation sont : 37°C, 5%CO₂ et 5% O₂.

Immunofluorescence 53BP1/Cycline A

Préparation des cellules

1. Ensemencer les cellules
 2. Les traiter, ou non avec de l'aphidicoline 200nM pendant 24h ou 72h. Pour un traitement de 72h, renouveler le traitement avec de l'aphidicoline fraîche au bout de 36h.
 3. 15h avant de les récolter, éliminer l'aphidicoline par 2 lavages au milieu de culture à 37°C et les ensemencer de façon à ce qu'elles soient en phase exponentielle le lendemain.
 4. Rincer doucement les cellules dans du PBS à température ambiante.
 5. Fixer les cellules avec de la paraformaldehyde 4% dans du PBS 15 minutes à température ambiante.
 6. Faire 3 lavages au PBS.
- Les cellules peuvent également être stockées à 4°C à partir de cette étape.

Saturation et incubation avec les anticorps

1. Saturer les sites non-spécifiques en incubant 30 min dans de l'albumine de sérum bovin (BSA) à 5% dans du PBS.
2. Incuber avec les anticorps primaires dilués dans du PBS-BSA 1.5% durant 2 heures à température ambiante
3. Laver les cellules avec du PBS 3x pendant 5 minutes. Faire un dernier lavage avec du PBS-BSA 1.5%.
4. Incuber avec les anticorps secondaires couplés aux fluorochromes dilués au 1:1000 dans du PBS-BSA 1.5% durant 1h à température ambiante à l'obscurité.
5. Laver les cellules avec du PBS 3x pendant 5 min. Rester à l'obscurité.
6. Monter les lamelles dans du milieu de montage contenant du DAPI.
7. Laisser sécher sur la nuit.
8. Observer les lamelles au microscope à fluorescence ou conserver à 4°C.

Immunofluorescence EdU/BrdU

Préparation des cellules

1. Ensemencer les cellules
 2. Les traiter, ou non avec de l'aphidicoline 200nM pendant 24h ou 72h. Pour un traitement de 72h, renouveler le traitement avec de l'aphidicoline fraîche au bout de 36h.
 3. Réaliser une incorporation d'EdU 10µM pendant 30min
 4. Eliminer l'aphidicoline par 2 lavages au milieu de culture à 37°C.
 - Pour le temps t0, réaliser un pulse BrdU 10mM pendant 30 min avant de les fixer comme décrit ensuite.
 - Pour les temps long de relargage (entre 9h et 15h), les ensemer de façon à ce qu'elles soient en phase exponentielle le lendemain.
 5. A chaque temps de récolte, procéder au pulse BrdU 10mM, 30min.
 6. Rincer doucement les cellules dans du PBS à température ambiante.
 7. Fixer les cellules avec de la paraformaldehyde 4% dans du PBS 15 minutes à température ambiante.
 6. Faire 3 lavages au PBS.
- Les cellules peuvent également être stockées à 4°C à partir de cette étape.

Dénaturation de l'ADN

1. Incuber les cellules 30min en présence d'HCl 2N à température ambiante.
2. Faire 5 lavages au PBS

Saturation, Click-it EdUet incubation avec les anticorps

1. Réaliser la réaction de ClickIt selon les instructions du fabricant. Rester à l'obscurité toute la suite de l'expérience.
2. Saturer les sites non-spécifiques en incubant 30 min dans de l'albumine de sérum bovin (BSA) à 5% dans du PBS.
3. Incuber avec les anticorps primaires dilués dans du PBS-BSA 1.5% durant 2 heures à température ambiante
4. Laver les cellules avec du PBS 3x pendant 5 minutes. Faire un dernier lavage avec du PBS-BSA 1.5%.
5. Incuber avec les anticorps secondaires couplés aux fluorochromes dilués au 1:1000 dans du PBS-BSA 1.5% durant 1h à température ambiante.

6. Laver les cellules avec du PBS 3x pendant 5 min.
7. Incuber les lamelles 15min avec du DAPI 10mM dilué au 2500^{ème} dans du PBS.
8. Faire 2 lavages de 10min au PBS
9. Monter les lamelles
10. Laisser sécher sur la nuit
11. Observer les lamelles au microscope à fluorescence.
12. Analyse des résultats par *Cell profiler*.

Proximity Ligation Assay (PLA)

Préparation des cellules

1. La veille de les récolter,ensemencer les cellules sur des lames munies de 8 chambres de culture.
 2. Rincer doucement les cellules dans du PBS à température ambiante.
 3. Fixer les cellules avec de la paraformaldehyde 4% dans du PBS 15 minutes à température ambiante.
 4. Faire 3 lavages au PBS
- Les cellules peuvent être stockées à 4°C à partir de cette étape dans le PBS.

Saturation, Klik-it EdUet incubation avec les anticorps

1. Réaliser l'étape de saturation avec le tampon de blocage (50µL/chambre) fourni dans le kit pendant 30 minutes à 37°C dans une atmosphère humide.
2. Oter les chambres de cultures
3. Déposer sur les cellules 40µl d'anticorps primaires dilués dans le tampon de dilution inclus dans le kit. Incuber pendant 1h30 à 37 °C dans une atmosphère humide.
4. Laver les lames 2 fois au tampon A (Tris pH=7,4 0.01M, NaCl 0,15M, Tween20 0.05%) pendant 5 minutes sous agitation.
5. Incuber les lames en présence d'un mélange des sondes « + » et « - » diluées au cinquième dans le tampon de dilution inclus dans le kit à 37°C sous atmosphère humide pendant 1 heure.
6. Faire deux lavages de 5 minutes au tampon A sous agitation.
7. Incuber les lames avec la ligase diluée dans le tampon de ligation pendant 30 minutes à 37°C dans une atmosphère humide.
8. Faire deux lavages de 5 minutes au tampon A sous agitation.

9. Incuber les cellules avec la polymérase diluée dans le tampon d'amplification contenant les dNTP fluorescents pendant 1h40 à 37°C dans une atmosphère humide.
10. Faire un lavage au tampon B 1X avec DAPI 4µM pendant 10 minutes
11. Faire un lavage au tampon B 1X pendant 10 minutes
12. Faire un lavage au tampon B 0.01X
13. Monter les lames
14. Observer les lames au microscope.

Immuno-précipitations

Préparation des cellules

1. Ensemencer les cellules.
2. Traiter les cellules à l'aphidicoline 200nM pendant 72h pour la condition traitée, ou au DMSO 1/1000^{ème} pour la condition non traitée. Penser à rafraichir le milieu au bout de 36h.
3. Laver les cellules par du milieu complet et traiter au nocodazole 0.1µM pendant 17h.
4. Réaliser un « *shake-off* » mitotique pour récupérer les cellules en mitose.
5. Faire 3 lavages des cellules par du milieu complet à 37°C.
6. Les remettre en culture et incubé 3h.
7. Récupérer les cellules
8. Faire 2 rinçages PBS froid
9. Congeler le culot par « *snap freeze* » dans l'azote liquide

Préparation des extraits nucléaires

Travailler à 4°C ou dans la glace tout le temps

Isoler les noyaux

1. Utiliser les graduations pour mesurer le culot
2. Rapidement resuspendre le culot dans du tampon hypotonique avec 5x le volume du culot. Centrifuger avec le rotor JS-4.2 5min à 3000rpm.
3. Resuspendre les cellules dans le tampon hypotonique avec 3x le volume du culot initial, et laisser gonfler 10 min.
4. Transférer les cellules dans un « *dounce homogenizer* ». Homogénéiser avec 10 aller-retours en utilisant le piston de type B.

5. Transférer les cellules dans des tubes à centrifuger. Collecter les noyaux en centrifugeant 15 min à 4000rpm avec le rotor JS4.2. Le surnageant est la fraction cytoplasmique.

Extraction des noyaux

1. En utilisant les graduations sur les tubes, mesurer le volume du culot, ou « *packed nuclear volume* » (pnv) de l'étape précédente. Resuspendre les noyaux dans ½ pnv avec le tampon « *low salt* ».

2. Aux goutte-à-goutte, pendant une agitation ajouter ½ pnv de tampon « *high salt* ». Homogénéiser au *dounce* si nécessaire.

3. Laisser les noyaux 30 min avec une constante agitation. L'agitation doit être douce et lente (ex : balancelle, ou éventuellement roue très doucement)

4. Culotter les noyaux extraits en centrifugeant 30min à 14500rpm avec le rotor JA 20. Le surnageant est l'extrait nucléaire.

Dans la même centrifugation, clarifier la fraction cytoplasmique.

Dialyse et conservation de l'extrait

1. Placer l'extrait nucléaire dans un tube de dialyse, et dialyser contre 50 volumes de tampon de dialyse pendant 1h à 2h à 4°C.

2. Enlever l'extrait du tube de dialyse. Centrifuger l'extrait 20 min à 14500rpm avec le rotor JA-20. Jeter le culot.

3. Doser les protéines

4. Congeler par « *snap freeze* » et stocker à -80°C.

Tampons :

Tampon Hypotonique :

- 10mM Hepes pH 7.9 4°C
- 1.5mM MgCl₂
- 10mM KCl
- 0,5mM DTT
- PMSF 0.2mM

Tampon Low Salt :

- 20mM Hepes pH 7.9
- 25% glycérol
- 1.5mM MgCl₂

- 0.02M KCl
- 0.2mM EDTA
- 0.5mM DTT
- PMSF 0.2mM

Tampon High Salt :

- 20mM Hepes pH 7.9
- 25% glycérol
- 1.5mM MgCl₂
- 1.2M KCl
- 0.2mM EDTA
- 0.5mM DTT
- PMSF 0.2mM

Tampon de dialyse :

- 20 mM Hepes pH 7.9
- 20% Glycérol
- 100mM KCl
- 0.2mM EDTA
- 0.5mM DTT
- PMSF 0.2mM

Immunoprécipitation

Protocole pour 1 mg de protéines

1. Décongeler l'extrait nucléaire doucement sur glace.
2. Préparer les billes magnétiques : compter 60µl de billes par condition, en mélangeant les Dynabeads prot A et prot G volume à volume, soit un volume total V.
 - Faire 2 lavages dans 3xV de PBS, triton 0.05%
 - Faire 1 lavage dans 3xV de PBS
 - Reprendre dans le volume V de tampon de dialyse.
3. Réaliser un pré-clearing des extraits : Ajouter 30µl de billes dans l'extrait. Incuber 1h à 4°C, sur roue. Récupérer le surnageant qui correspond à l'*input*. En prélever un aliquot et réserver.
4. Sur le reste de l'extrait, ajouter 1µl de benzonase et de MgCl₂ pour une concentration finale de 2mM.
5. Ajouter 3µg d'anticorps anti-POLθ ou IgG contrôle à partir d'une solution à 1µg/µl. Incuber 2h30 sur roue à 4°C.
6. Ajouter 30 µl de billes par tube. Incuber 2h30 sur roue à 4°C.

7. Récupérer les billes. Le surnageant correspond au surnageant d'IP. En conserver un aliquot.
8. Faire 3 lavages de 5 min des billes au PBS
9. Reprendre les billes dans 100 à 150µl de peptide qui a servi à produire l'anticorps à 300µg/ml dilué dans de l'ammonium bicarbonate 500mM.
10. Incuber sur la nuit sur roue à vitesse très lente.
11. Récupérer le surnageant pour l'analyser par Western Blot

Cytométrie en Flux

Préparation des cellules

1. Au moment souhaité, incuber les cellules avec du BrdU 10µM pendant 15min à 37°C
 2. Faire un lavage à la thymidine à 30µM dans milieu complet
 3. Récupérer les cellules et faire un lavage au PBS
 4. Fixer les cellules à l'éthanol 70% à -20°C sur la nuit.
- A cette étape, les cellules peuvent être conservées à -20°C jusqu'à utilisation.

Marquage BrdU et analyse du cycle cellulaire par FACS :

1. Centrifuger les cellules 5min, 1000rpm et jeter le surnageant
2. Faire un lavage au PBS- 0,5% BSA.
3. Incuber 20min dans l'HCl 2N
4. Faire un lavage au PBS- 0,5% BSA.
5. Incuber 2min avec du Borate de Sodium 0,1M, pH=8,5
6. Faire un lavage au PBS- 0,5% BSA.
7. Incuber avec l'anticorps primaire anti-BrdU (1:50 dans PBS-0,5% BSA-0,1% triton) pendant 1h à température ambiante.
8. Faire un lavage au PBS- 0,5% BSA.
9. Incuber avec l'anticorps secondaire couplé à l'Alexa 488 (1:250 dans PBS-0,5% BSA-0,1% triton) pendant 30min, à température ambiante, à l'obscurité.
10. Faire un lavage au PBS
11. Reprendre dans du PBS contenant de la RNase (100 mg/ml) et de l'iodure de propidium à 10µg/ml. Incuber 30min sous agitation avant analyse par FACS.

Halos d'ADN

Préparation des cellules

1. Ensemencer les cellules RKO
2. Transfecter par 50nM de siRNA *CTL*, *POLQ-1* ou *POLQ-2*, à la lipofectamine 2000.
3. 32h après transfection, traiter les cellules au nocodazole 0,1μM, incubé 21h
4. Collecter les cellules en mitose par *shake-off* mitotique
5. Faire 2 lavages au milieu complet chaud
6. Remettre en culture pendant 3h.
7. Traiter les cellules à la L-mimosine 0,4μM pendant 15h
8. Récolter les cellules

Préparation des noyaux

1. Laver les cellules au PBS.
2. Reprendre le culot cellulaire dans le « tampon nucléaire », 0.5% NP40
Incuber sur glace pendant 5 min
3. Centrifuger à 600g, 5min, 4°C
4. Faire un lavage au tampon nucléaire
5. Reprendre le culot dans le tampon nucléaire (200000 cells/ml)
6. Centrifuger au cytospin, 1500rpm
7. Laisser sécher la lamelle
8. Faire le marquage au DAPI 2μg/ml dans PBS pendant 4min
9. Incuber 1 min dans le « tampon NaCl »
10. Incuber 4min dans le « tampon Halo »
11. Incuber 30s dans le tampon de lavage 1
12. Incuber 30s dans le tampon de lavage 2
13. Incuber 10min dans la formaldéhyde 2% (dans H₂O)
14. Monter les lames
15. Observation au microscope à fluorescence
16. Le rayon (R) du noyau a été obtenu en mesurant l'aire totale du noyau (A_t) et l'aire de la partie centrale très intense en DAPI (A_c), et en utilisant la formule suivante:

$$R : \sqrt[3]{(A_t/\pi)} - \sqrt[3]{(A_c/\pi)}.$$

Composition des tampons

Tampon nucléaire :

- 10mM Tris-HCl pH 8
- 100mM NaCl
- 0,3M sucrose
- 3mM MgCl₂
- Inhibiteurs protéases et phosphatases

Tampon NaCl :

- 25mM Tris-HCl pH 8
- 0,5M NaCl
- 0,2mM MgCl₂
- 1mM

Tampon Halo :

- 10mM Tris-HCl pH 8
- 10mM EDTA
- 2M NaCl
- 1mM DTT
- Inhibiteurs protéases et phosphatases

Tampon de lavage 1 :

- 0,2M NaCl
- 0,2mM MgCl₂
- 25mM Tris-HCl pH 8

Tampon de lavage 2 :

- 0,2mM MgCl₂
- 25mM Tris-HCl pH 8

BIBLIOGRAPHIE

- Achkar, E.E., and Gerbault-Seureau, M (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proc Natl Acad Sci U S A*;102(50):18069-74
- Alabert, C., and Groth, A. (2012). Chromatin replication and epigenome maintenance. *Nat. Rev. Mol. Cell Biol.* 13, 153–167.
- Allera-Moreau, C., Rouquette, I., Lepage, B., Oumouhou, N., Walschaerts, M., Leconte, E., Schilling, V., Gordien, K., Brouchet, L., Delisle, M.B., et al. (2012). DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis* 1, e30.
- Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M., et al. (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* 151, 835–846.
- Amiel, A, Litmanovitch, T, and Lishner, M (1998a). Temporal differences in replication timing of homologous loci in malignant cells derived from CML and lymphoma patients. *Genes Chromosomes Cancer.* 22(3):225-31.
- Amiel, A, Kirgner, I, Gaber, E, Manor, Y, and Fejgin, M (1999). Replication pattern in cancer: asynchronous replication in multiple myeloma and in monoclonal gammopathy. *Cancer Genet Cytogenet.* 108(1):32-7.
- Amiel, A., Kolodizner, T., Fishman, A., Gaber, E., Klein, Z., Beyth, Y., and Fejgin, M.D. (1998b). Replication pattern of the p53 and 21q22 loci in the premalignant and malignant stages of carcinoma of the cervix. *Cancer* 83, 1966–1971.
- Anglana, M, Apiou, F, Bensimon, A, and Debatisse, M (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* 114(3):385-94.
- Arana, M.E., Seki, M., Wood, R.D., Rogozin, I.B., and Kunkel, T.A. (2008). Low-fidelity DNA synthesis by human DNA polymerase theta. *Nucleic Acids Res.* 36, 3847–3856.
- Arias, E.E., and Walter, J.C. (2006). PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication. *Nat. Cell Biol.* 8, 84–90.
- Asagoshi, K., Lehmann, W., Braithwaite, E.K., Santana-Santos, L., Prasad, R., Freedman, J.H., Van Houten, B., and Wilson, S.H. (2012). Single-nucleotide base excision repair DNA polymerase activity in *C. elegans* in the absence of DNA polymerase β . *Nucleic Acids Res.* 40, 670–681.
- Babushok, D.V., and Kazazian, H.H. (2007). Progress in understanding the biology of the human mutagen LINE-1. *Hum. Mutat.* 28, 527–539.
- Baddeley, D., Chagin, V.O., Schermelleh, L., Martin, S., Pombo, A., Carlton, P.M., Gahl, A., Domaing, P., Birk, U., Leonhardt, H., et al. (2010). Measurement of replication structures at the nanometer scale using super-resolution light microscopy. *Nucleic Acids Research* 38, e8.
- Barbosa, A., Otto, P.A., and Vianna-Morgante, A.M. (2000). Replication timing of homologous -satellite DNA in Roberts syndrome. *Chromosome Research* 8, 645–650.
- Bartek, J., Bartkova, J., and Lukas, J. (1996). The retinoblastoma protein pathway and the restriction point. *Curr. Opin. Cell Biol.* 8, 805–814.
- Bartek, J., Lukas, J., and Bartkova, J. (2007). DNA damage response as an anti-cancer barrier: damage threshold and the concept of “conditional haploinsufficiency”. *Cell Cycle* 6, 2344–2347.

- Bartkova, J., Horejsí, Z., Koed, K., Krämer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M., Lukas, C., et al. (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 434, 864–870.
- Beck, H., Nähse-Kumpf, V., Larsen, M.S., O’Hanlon, K.A., Patzke, S., Holmberg, C., Mejlvang, J., Groth, A., Nielsen, O., Syljuåsen, R.G., et al. (2012). Cyclin-dependent kinase suppression by WEE1 kinase protects the genome through control of replication initiation and nucleotide consumption. *Mol. Cell. Biol.* 32, 4226–4236.
- Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schübeler, D. (2010). Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nature Structural & Molecular Biology* 17, 894–900.
- Belsky, J.A., MacAlpine, H.K., Lubelsky, Y., Hartemink, A.J., and MacAlpine, D.M. (2015). Genome-wide chromatin footprinting reveals changes in replication origin architecture induced by pre-RC assembly. *Genes Dev.* 29, 212–224.
- Berbenetz, N.M., Nislow, C., and Brown, G.W. (2010). Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet.* 6, e1001092.
- Bergoglio, V., Boyer, A.-S.S., Walsh, E., Naim, V., Legube, G., Lee, M.Y., Rey, L., Rosselli, F., Cazaux, C., Eckert, K.A., et al. (2013). DNA synthesis by Pol η promotes fragile site stability by preventing under-replicated DNA in mitosis. *J. Cell Biol.* 201, 395–408.
- Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. *Gene* 241, 3–17.
- Besnard, E., Babled, A., Lapasset, L., Milhavel, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Structural & Molecular Biology* 19, 837–844.
- Bester, A.C., Roniger, M., Oren, Y.S., Im, M.M., Sarni, D., Chaoat, M., Bensimon, A., Zamir, G., Shewach, D.S., and Kerem, B. (2011). Nucleotide deficiency promotes genomic instability in early stages of cancer development. *Cell* 145, 435–446.
- Bhat, A., Andersen, P.L., Qin, Z., and Xiao, W. (2013). Rev3, the catalytic subunit of Pol ζ , is required for maintaining fragile site stability in human cells. *Nucleic Acids Research.* 41(4):2328-39
- Botuyan, M., Lee, J., Ward, I.M., Kim, J.-E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127, 1361–1373.
- Boyer, A.-S.S., Grgurevic, S., Cazaux, C., and Hoffmann, J.-S.S. (2013). The human specialized DNA polymerases and non-B DNA: vital relationships to preserve genome integrity. *J. Mol. Biol.* 425, 4767–4781.
- Breger, K.S., Smith, L., Turker, M.S., and Thayer, M.J. (2004). Ionizing radiation induces frequent translocations with delayed replication and condensation. *Cancer Research.* 64(22):8231-8
- Bunting, S.F., Callén, E., Wong, N., Chen, H.-T.T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., et al. (2010). 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243–254.
- Buongiorno-Nardelli, M., Micheli, G., Carri, M.T., and Marilley, M. (1982). A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature* 298, 100–102.

- Bétous, R., Rey, L., Wang, G., Pillaire, M.-J.J., Puget, N., Selves, J., Biard, D.S., Shin-ya, K., Vasquez, K.M., Cazaux, C., et al. (2009). Role of TLS DNA polymerases eta and kappa in processing naturally occurring structured DNA in human cells. *Mol. Carcinog.* *48*, 369–378.
- Bétous, R., Pillaire, M.-J.J., Pierini, L., van der Laan, S., Recolin, B., Ohl-Séguy, E., Guo, C., Niimi, N., Grúz, P., Nohmi, T., et al. (2013). DNA polymerase κ -dependent DNA synthesis at stalled replication forks is important for CHK1 activation. *EMBO J.* *32*, 2172–2185.
- Cadoret, J.-C.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 15837–15842.
- Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Research* *21*, 1438–1449.
- Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Méchali, M. (2012). New insights into replication origin characteristics in metazoans. *Cell Cycle (Georgetown, Tex.)* *11*, 658–667.
- Cea, V., Cipolla, L., and Sabbioneda, S. (2015). Replication of Structured DNA and its implication in epigenetic stability. *Front Genet* *6*, 209.
- Ceccaldi, R., Liu, J.C., Amunugama, R., Hajdu, I., Primack, B., Petalcorin, M.I., O'Connor, K.W., Konstantinopoulos, P.A., Elledge, S.J., Boulton, S.J., et al. (2015a). Homologous-recombination-deficient tumours are dependent on Pol θ -mediated repair. *Nature* *518*, 258–262.
- Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2015b). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* pii: S0962-8924(15)00142-7
- Chabosseau, P., Buhagiar-Labarchède, G., Onclercq-Delic, R., Lambert, S., Debatisse, M., Brison, O., and Amor-Guérét, M. (2011). Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. *Nat Commun* *2*, 368.
- Chan, K.-L.L., North, P.S., and Hickson, I.D. (2007). BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. *EMBO J.* *26*, 3397–3409.
- Chan, K.L., Palma-Pallag, T., Ying, S., and Hickson, I.D. (2009). Replication stress induces sister-chromatid bridging at fragile site loci in mitosis. *Nat. Cell Biol.* *11*, 753–760.
- Chan, S.H., Yu, A.M., and McVey, M. (2010). Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* *6*, e1001005.
- Chapman, JR, Barral, P, Vannier, JB, Borel, V, and Steger, M (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Molecular Cell.* *49*(5):858-71
- Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* *40*, 179–204.
- Clijsters, L., Ogink, J., and Wolthuis, R. (2013). The spindle checkpoint, APC/C(Cdc20), and APC/C(Cdh1) play distinct roles in connecting mitosis to S phase. *J. Cell Biol.* *201*, 1013–1026.

- Comoglio, F., Schlumpf, T., Schmid, V., Rohs, R., Beisel, C., and Paro, R. (2015). High-resolution profiling of *Drosophila* replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Reports* *11*, 821–834.
- Cornacchia, D., Dileep, V., Quivy, J.-P.P., Foti, R., Tili, F., Santarella-Mellwig, R., Antony, C., Almouzni, G., Gilbert, D.M., and Bonomo, S.B. (2012). Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *The EMBO Journal* *31*, 3678–3690.
- Costantini, M., and Bernardi, G. (2008). Replication timing, chromosomal bands, and isochores. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 3433–3437.
- Costantini, M., Cammarano, R., and Bernardi, G. (2009). The evolution of isochore patterns in vertebrate genomes. *BMC Genomics* *10*, 146.
- Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* *455*, 557–560.
- Cseresnyes, Z., Schwarz, U., and Green, C.M. (2009). Analysis of replication factories in human cells by super-resolution light microscopy. *BMC Cell Biology* *10*, 88.
- Cytron, S., Stepnov, E., Bounkin, I., and Mashevich, M. (2011). Epigenetic analyses in blood cells of men suspected of prostate cancer predict the outcome of biopsy better than serum PSA levels. *Clin Epigenetics* *2*(2):383-388.
- Das, S., Borrmann, T., Liu, V., Yang, S., Bechhoefer, J., and Rhind, N. (2015). Replication timing is regulated by the number of MCMs loaded at origins. *Genome Research*. *In press*
- Davé, A., Cooley, C., Garg, M., and Bianchi, A. (2014). Protein phosphatase 1 recruitment by Rif1 regulates DNA replication origin firing by counteracting DDK activity. *Cell Reports* *7*, 53–61.
- Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B., and Brison, O. (2012). Common fragile sites: mechanisms of instability revisited. *Trends Genet.* *28*, 22–32.
- Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res.* *23*, 1–11.
- Depamphilis, M. (1993). Origins of DNA replication that function in eukaryotic cells. *Current Opinion in Cell Biology* *5*, 434–441.
- DePamphilis, M.L. (1993). Eukaryotic DNA replication: anatomy of an origin. *Annual Review of Biochemistry* *62*, 29–63.
- Desprat, R., Thierry-Mieg, D., Lailier, N., Lajugie, J., Schildkraut, C., Thierry-Mieg, J., and Bouhassira, E.E. (2010). Predictable dynamic program of timing of DNA replication in human cells. *Genome Research* *19*, 2288–2299.
- Dhar, M., Sehgal, S., and Kaul, S. (2012). Structure, replication efficiency and fragility of yeast ARS elements. *Research in Microbiology* *163*, 243–253.
- Dimitrova, D.S., and Gilbert, D.M. (2000). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Molecular Cell* *4*, 983–993.
- Donley, N., and Thayer, M.J. (2013). DNA replication timing, genome stability and cancer: late and/or delayed DNA replication timing is associated with increased genomic instability. *Semin. Cancer Biol.* *23*, 80–89.
- Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. *Annu. Rev. Genet.* *41*:169-92

- Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. *Genes Dev.* 24, 748–753.
- Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T., Tkáč, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D., et al. (2013). A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell* 49, 872–883.
- Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. (2008). Global organization of replication time zones of the mouse genome. *Genome Research* 18, 1562–1570.
- Feng, L., Fong, K.-W., Wang, J., Wang, W., and Chen, J. (2013). RIF1 counteracts BRCA1-mediated end resection during DNA repair. *Journal of Biological Chemistry* 288, 11135–11143.
- Fernandez-Vidal, A., Guitton-Sert, L., Cadoret, J.-C.C., Drac, M., Schwob, E., Baldacci, G., Cazaux, C., and Hoffmann, J.-S.S. (2014). A role for DNA polymerase θ in the timing of DNA replication. *Nat Commun* 5, 4285.
- Foulk, M.S., Urban, J.M., Casella, C., and Gerbi, S.A. (2015). Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res.* 25, 725–735.
- Fradet-Turcotte, A., Canny, M.D., Escribano-Díaz, C., Orthwein, A., Leung, C.C., Huang, H., Landry, M.-C.C., Kitevski-LeBlanc, J., Noordermeer, S.M., Sicheri, F., et al. (2013). 53BP1 is a reader of the DNA-damage-induced H2A Lys 15 ubiquitin mark. *Nature* 499, 50–54.
- Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. *Nat. Rev. Mol. Cell Biol.* 16, 360–374.
- Friedberg, E.C. (2005). Suffering in silence: the tolerance of DNA damage. *Nat. Rev. Mol. Cell Biol.* 6, 943–953.
- Fritz, A., Sinha, S., and Marella, N (2013). Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. *J Cell Biochem.* 114(5):1074-83.
- Gambus, A., Khoudoli, G.A., Jones, R.C., and Blow, J.J. (2011). MCM2-7 form double hexamers at licensed origins in *Xenopus* egg extract. *J. Biol. Chem.* 286, 11855–11864.
- Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* 25, 2041–2056.
- Garg, P., and Burgers, P.M. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.* 40, 115–128.
- Gelot, C., Magdalou, I., and Lopez, B.S. (2015). Replication stress in Mammalian cells and its consequences for mitosis. *Genes (Basel)* 6, 267–298.
- Gemble, S., Ahuja, A., Buhagiar-Labarchède, G., Onclercq-Delic, R., Dairou, J., Biard, D.S., Lambert, S., Lopes, M., and Amor-Guérét, M. (2015). Pyrimidine Pool Disequilibrium Induced by a Cytidine Deaminase Deficiency Inhibits PARP-1 Activity, Leading to the Under Replication of DNA. *PLoS Genet.* 11, e1005384.
- Gilbert, D.M. (1986). Temporal order of replication of *Xenopus laevis* 5S ribosomal RNA genes in somatic cells. *Proceedings of the National Academy of Sciences of the United States of America* 83, 2924–2928.
- Gilbert, D.M. (2007). Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. *Chromosoma* 116, 341–347.

- Gindin, Y., Valenzuela, M.S., Aladjem, M.I., Meltzer, P.S., and Bilke, S. (2014). A chromatin structure-based model accurately predicts DNA replication timing in human cells. *Molecular Systems Biology* 10, 722.
- Goff, J.P., Shields, D.S., Seki, M., Choi, S., Epperly, M.W., Dixon, T., Wang, H., Bakkenist, C.J., Dertinger, S.D., Torous, D.K., et al. (2009). Lack of DNA polymerase theta (POLQ) radiosensitizes bone marrow stromal cells in vitro and increases reticulocyte micronuclei after total-body irradiation. *Radiat. Res.* 172, 165–174.
- Gorgoulis, V.G., Vassiliou, L.-V.F.V., Karakaidos, P., Zacharatos, P., Kotsinas, A., Liloglou, T., Venere, M., Ditullio, R.A., Kastrinakis, N.G., Levy, B., et al. (2005). Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 434, 907–913.
- Greally, JM, Cuker, A, and Bowers, PN (2003). Epigenetic abnormalities associated with a chromosome 18 (q21-q22) inversion and a Gilles de la Tourette syndrome phenotype. *Proc Natl Acad Sci U S A.* 100(8):4684-9.
- Guan, Z., Hughes, C.M., Kosiyatrakul, S., Norio, P., Sen, R., Fiering, S., Allis, C.D., Bouhassira, E.E., and Schildkraut, C.L. (2009). Decreased replication origin activity in temporal transition regions. *The Journal of Cell Biology* 187, 623–635.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951.
- Guilbaud, G., Rappailles, A., Baker, A., Chen, C.-L.L., Arneodo, A., Goldar, A., d’Aubenton-Carafa, Y., Thermes, C., Audit, B., and Hyrien, O. (2012). Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Computational Biology* 7, e1002322.
- Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., and Méndez, J. (2011). Cohesin organizes chromatin loops at DNA replication factories. *Genes & Development* 24, 2812–2822.
- Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. *Science* 319, 1352–1355.
- Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences of the United States of America* 107, 139–144.
- Harrigan, J.A., Belotserkovskaya, R., Coates, J., Dimitrova, D.S., Polo, S.E., Bradshaw, C.R., Fraser, P., and Jackson, S.P. (2011). Replication stress induces 53BP1-containing OPT domains in G1 cells. *J. Cell Biol.* 193, 97–108.
- Hassan-Zadeh, V., Chilaka, S., Cadoret, J.-C.C., Ma, M.K., Boggetto, N., West, A.G., and Prioleau, M.-N.N. (2012). USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS Biol.* 10, e1001277.
- Hayano, M., Kanoh, Y., Matsumoto, S., Renard-Guillet, C., Shirahige, K., and Masai, H. (2012). Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes & Development* 26, 137–150.
- Heard, E., and Disteche, C.M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 20, 1848–1867.
- Heichinger, C., Penkett, C.J., Bähler, J., and Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. *The EMBO Journal* 25, 5171–5179.

- Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E., and Heiden, T. (2006). Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res.* 16, 1222–1230.
- Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* 44, 966–977.
- Hengeveld, R., Boer, H., Schoonen, P., Vries, E., Lens, S., and Vugt, M. (2015). Rif1 Is Required for Resolution of Ultrafine DNA Bridges in Anaphase to Ensure Genomic Stability. *Developmental Cell* 34.
- Higgins, G.S., Prevo, R., Lee, Y.-F.F., Helleday, T., Muschel, R.J., Taylor, S., Yoshimura, M., Hickson, I.D., Bernhard, E.J., and McKenna, W.G. (2010a). A small interfering RNA screen of genes involved in DNA repair identifies tumor-specific radiosensitization by POLQ knockdown. *Cancer Res.* 70, 2984–2993.
- Higgins, G.S., Harris, A.L., Prevo, R., Helleday, T., McKenna, W.G., and Buffa, F.M. (2010b). Overexpression of POLQ confers a poor prognosis in early breast cancer patients. *Oncotarget* 1, 175–184.
- Hiraga, S.-I., Alvino, G.M., Chang, F., Lian, H.-Y.Y., Sridhar, A., Kubota, T., Brewer, B.J., Weinreich, M., Raghuraman, M.K., and Donaldson, A.D. (2014). Rif1 controls DNA replication by directing Protein Phosphatase 1 to reverse Cdc7-mediated phosphorylation of the MCM complex. *Genes & Development* 28, 372–383.
- Hiratani, I., and Gilbert, D.M. (2009). Replication timing as an epigenetic mark. *Epigenetics* 4, 93–97.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T., Schübeler, D., and Gilbert, D. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. *PLoS Biology*. 6(10):e245
- Hogg, M., Seki, M., Wood, R.D., Doublié, S., and Wallace, S.S. (2011). Lesion bypass activity of DNA polymerase θ (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. *J. Mol. Biol.* 405, 642–652.
- Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J. Biol. Chem.* 288, 30161–30171.
- Huppert, J.L. (2010). Structure, location and interactions of G-quadruplexes. *FEBS J.* 277, 3452–3458.
- Hyrien, O., Maric, C., and Méchali, M. (1995). Transition in specification of embryonic metazoan DNA replication origins. *Science (New York, N.Y.)* 270, 994–997.
- Ibarra, A., Schwob, E., and Méndez, J. (2008). Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proceedings of the National Academy of Sciences of the United States of America* 105, 8956–8961.
- Im, J.-S.S., Ki, S.-H.H., Farina, A., Jung, D.-S.S., Hurwitz, J., and Lee, J.-K.K. (2009). Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15628–15632.
- Imai, K., and Yamamoto, H. (2008). Carcinogenesis and microsatellite instability: the interrelationship between genetics and epigenetics. *Carcinogenesis* 29, 673–680.
- Inagaki, S., Suzuki, T., Ohto, M.A., Urawa, H., Horiuchi, T., Nakamura, K., and Morikami, A. (2006). Arabidopsis TEBICHI, with helicase and DNA polymerase domains, is required for regulated cell division and differentiation in meristems. *Plant Cell* 18, 879–892.

- Jasencakova, Z, Scharf, A., Ask, K, Corpet, A, and Imhof, A (2010). Replication stress interferes with histone recycling and predeposition marking of new histones. *Mol Cell* 37(5):736-43
- Johnson, R.E., Klassen, R., Prakash, L., and Prakash, S. (2015). A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands. *Mol. Cell* 59, 163–175.
- Jones, RM, Mortusewicz, O, Afzal, I, and Lorvellec, M (2013). Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* 32(32):3744-53
- Julienne, H., Zoufir, A., Audit, B., and Arneodo, A. (2013). Human genome replication proceeds through four chromatin states. *PLoS Computational Biology* 9, e1003233.
- Kanoh, J., and Ishikawa, F. (2001). spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Curr. Biol.* 11, 1624–1630.
- Kanoh, Y., Matsumoto, S., Fukatsu, R., Kakusho, N., Kono, N., Renard-Guillet, C., Masuda, K., Iida, K., Nagasawa, K., Shirahige, K., et al. (2015). Rif1 binds to G quadruplexes and suppresses replication over long distances. *Nat. Struct. Mol. Biol.* *In press*
- Katsuno, Y., Suzuki, A., Sugimura, K., Okumura, K., Zineldeen, D.H., Shimada, M., Niida, H., Mizuno, T., Hanaoka, F., and Nakanishi, M. (2009). Cyclin A-Cdk1 regulates the origin firing program in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* 106, 3184–3189.
- Kawamura, K., Bahar, R., Seimiya, M., Chiyo, M., Wada, A., Okada, S., Hatano, M., Tokuhisa, T., Kimura, H., Watanabe, S., et al. (2004). DNA polymerase theta is preferentially expressed in lymphoid tissues and upregulated in human cancers. *Int. J. Cancer* 109, 9–16.
- Kent, T., Chandramouly, G., McDevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase θ . *Nat. Struct. Mol. Biol.* 22, 230–237.
- Khurana, S., and Oberdoerffer, P. (2015). Replication Stress: A Lifetime of Epigenetic Change. *Genes (Basel)* 6, 858–877.
- Kim, J.C., and Mirkin, S.M. (2013). The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* 23, 280–288.
- Koole, W., van Schendel, R., Karambelas, A., van Heteren, J., Okihara, K., and Tijsterman, M. (2014). A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun* 5:3216.
- Korenstein-Ilan, A, Amiel, A, Lalezari, S, and Lishner, M (2002). Allele-specific replication associated with aneuploidy in blood cells of patients with hematologic malignancies. *Cancer Genet Cytogenet.* 139(2):97-103.
- Koundrioukoff, S., Carignon, S., Técher, H., Letessier, A., Brison, O., and Debatisse, M. (2013). Stepwise activation of the ATR signaling pathway upon increasing replication stress impacts fragile site integrity. *PLoS Genet.* 9, e1003643.
- Kumagai, A., Shevchenko, A., Shevchenko, A., and Dunphy, W.G. (2010). Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell* 140, 349–359.
- Kunkel, T.A. (2011). Balancing eukaryotic replication asymmetry with replication fidelity. *Curr Opin Chem Biol* 15, 620–626.
- Labib, K., Diffley, J.F., and Kearsley, S.E. (1999). G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nat. Cell Biol.* 1, 415–422.
- Lang, GI, and Murray, AW (2011). Mutation rates across budding yeast chromosome VI are correlated with replication timing. *Genome Biology and Evolution.* 3:799-811.

- Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J., and Bensimon, A. (2007). DNA replication origin interference increases the spacing between initiation events in human cells. *Molecular Biology of the Cell* 17, 5337–5345.
- Lee, D.-H., Acharya, S., Kwon, M., Drane, P., Guan, Y., Adelmant, G., Kalev, P., Shah, J., Pellman, D., Marto, J., et al. (2014). Dephosphorylation Enables the Recruitment of 53BP1 to Double-Strand DNA Breaks. *Molecular Cell* 54(3):512-25.
- Lemaitre, J.-M.M., Danis, E., Pasero, P., Vassetzky, Y., and Méchali, M. (2006). Mitotic remodeling of the replicon and chromosome structure. *Cell* 123, 787–801.
- Lemée, F., Bergoglio, V., Fernandez-Vidal, A., Machado-Silva, A., Pillaire, M.-J.J., Bieth, A., Gentil, C., Baker, L., Martin, A.-L.L., Leduc, C., et al. (2010). DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13390–13395.
- Letessier, A., Millot, G.A.A., Koundrioukoff, S., Lachagès, A.-M.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470, 120–123.
- Li, A., and Blow, J.J. (2005). Cdt1 downregulation by proteolysis and geminin inhibition prevents DNA re-replication in *Xenopus*. *The EMBO Journal*. 24(2):395-404
- Li, C.-J.J., and DePamphilis, M.L. (2002). Mammalian Orc1 protein is selectively released from chromatin and ubiquitinated during the S-to-M transition in the cell division cycle. *Mol. Cell. Biol.* 22, 105–116.
- Liu, E., Lee, A.Y., Chiba, T., Olson, E., Sun, P., and Wu, X. (2007). The ATR-mediated S phase checkpoint prevents rereplication in mammalian cells when licensing control is disrupted. *J. Cell Biol.* 179, 643–657.
- Liu, S., Bekker-Jensen, S., Mailand, N., Lukas, C., Bartek, J., and Lukas, J. (2006). Claspin operates downstream of TopBP1 to direct ATR signaling towards Chk1 activation. *Mol. Cell. Biol.* 26, 6056–6064.
- Liu, Y., Nielsen, C., Yao, Q., and Hickson, I. (2014). The origins and processing of ultra fine anaphase DNA bridges. *Curr. Opin. Genet. Dev.* 26, 1–5.
- Lombrana, R., Almeida, R., Revuelta, I., Madeira, S., Herranz, G., Saiz, N., Bastolla, U., and Gómez, M. (2013). High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *EMBO J.* 32, 2631–2644.
- Lombrana, R., Almeida, R., Álvarez, A., and Gómez, M. (2015). R-loops and initiation of DNA replication in human cells: a missing link? *Frontiers in Genetics* 6, 158.
- Lottersberger, F., and Bothmer, A. (2013). Role of 53BP1 oligomerization in regulating double-strand break repair. *Proc Natl Acad Sci U S A*. 110(6):2146-51..
- Loupart, M.L., Krause, S.A., and Heck, M.S. (2000). Aberrant replication timing induces defective chromosome condensation in *Drosophila* ORC2 mutants. *Curr. Biol.* 10, 1547–1556.
- Lu, J., Li, F., Murphy, C., Davidson, M., and Gilbert, D. (2010). G2 phase chromatin lacks determinants of replication timing. *The Journal of Cell Biology* 189.
- Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R.S., Grøfte, M., Chan, K.L., Hickson, I.D., Bartek, J., et al. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. *Nat. Cell Biol.* 13, 243–253.
- Ma, H., Samarabandu, J., Devdhar, R.S., Acharya, R., Cheng, P.C., Meng, C., and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. *J. Cell Biol.* 143, 1415–1425.

- MacAlpine, D.M., Rodríguez, H.K., and Bell, S.P. (2005). Coordination of replication and transcription along a *Drosophila* chromosome. *Genes & Development* 18, 3094–3105.
- MacAlpine, H.K., Gordân, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* 20, 201–211.
- Maga, G., Shevelev, I., Ramadan, K., Spadari, S., and Hübscher, U. (2002). DNA polymerase theta purified from human cells is a high-fidelity enzyme. *J. Mol. Biol.* 319, 359–369.
- Mailand, N., Bekker-Jensen, S., Bartek, J., and Lukas, J. (2006). Destruction of Claspin by SCFbetaTrCP restrains Chk1 activation and facilitates recovery from genotoxic stress. *Mol. Cell* 23, 307–318.
- Mantiero, D., Mackenzie, A., Donaldson, A., and Zegerman, P. (2011). Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *The EMBO Journal* 30, 4805–4814.
- Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. *Nature Reviews Genetics* 11, 285–296.
- Marheineke, K., and Hyrien, O. (2004). Control of replication origin density and firing time in *Xenopus* egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint. *The Journal of Biological Chemistry* 279, 28071–28081.
- Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Research* 21, 1822–1832.
- Martomo, S.A., Saribasak, H., Yokoi, M., Hanaoka, F., and Gearhart, P.J. (2008). Reevaluation of the role of DNA polymerase theta in somatic hypermutation of immunoglobulin genes. *DNA Repair (Amst.)* 7, 1603–1608.
- Masuda, K., Ouchida, R., Hikida, M., Nakayama, M., Ohara, O., Kurosaki, T., and O-Wang, J. (2006). Absence of DNA polymerase theta results in decreased somatic hypermutation frequency and altered mutation patterns in Ig genes. *DNA Repair (Amst.)* 5, 1384–1391.
- Masuda, K., Ouchida, R., Hikida, M., Kurosaki, T., Yokoi, M., Masutani, C., Seki, M., Wood, R.D., Hanaoka, F., and O-Wang, J. (2007). DNA polymerases eta and theta function in the same genetic pathway to generate mutations at A/T during somatic hypermutation of Ig genes. *J. Biol. Chem.* 282, 17387–17394.
- Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E., and Sfeir, A. (2015). Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* 518, 254–257.
- Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y., and Takisawa, H. (2006). The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol. Cell. Biol.* 26, 4843–4852.
- Maya-Mendoza, A., Petermann, E., Gillespie, D.A., Caldecott, K.W., and Jackson, D.A. (2007). Chk1 regulates the density of active replication origins during the vertebrate S phase. *The EMBO Journal* 26, 2719–2731.
- McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.* 18, 148–161.
- McGarry, T.J., and Kirschner, M.W. (1998). Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* 93, 1043–1053.
- Mejlvang, J., Feng, Y., Alabert, C., Neelsen, K.J., Jasencakova, Z., Zhao, X., Lees, M., Sandelin, A., Pasero, P., Lopes, M., et al. (2014). New histone supply regulates replication fork speed and PCNA unloading. *J. Cell Biol.* 204, 29–43.

- Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Research* 21, 377–389.
- Miotto, B., and Struhl, K. (2008). HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes & Development* 22, 2633–2638.
- Moses, A.M., Liku, M.E., Li, J.J., and Durbin, R. (2007). Regulatory evolution in proteins by turnover and lineage-specific changes of cyclin-dependent kinase consensus sites. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17713–17718.
- Méndez, J., and Stillman, B. (2000). Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* 20, 8602–8612.
- Méndez, J., Zou-Yang, X.H., Kim, S.-Y.Y., Hidaka, M., Tansey, W.P., and Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. *Mol. Cell* 9, 481–491.
- Naim, V., and Rosselli, F. (2009a). The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities. *Nat. Cell Biol.* 11, 761–768.
- Naim, V., and Rosselli, F. (2009b). The FANC pathway and mitosis: A replication legacy. *Cell Cycle* 8, 2907–2912.
- Naim, V., Wilhelm, T., Debatisse, M., and Rosselli, F. (2013). ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nat. Cell Biol.* 15, 1008–1015.
- Nakayasu, H., and Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. *The Journal of Cell Biology* 108, 1–11.
- Nam, E.A., and Cortez, D. (2011). ATR signalling: more than meeting at the fork. *Biochem. J.* 436, 527–536.
- Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* 9, 27–43.
- Nguyen, V.Q., Co, C., Irie, K., and Li, J.J. (2000). Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr. Biol.* 10, 195–205.
- Nishitani, H., Sugimoto, N., Roukos, V., Nakanishi, Y., Saijo, M., Obuse, C., Tsurimoto, T., Nakayama, K.I., Nakayama, K., Fujita, M., et al. (2006). Two E3 ubiquitin ligases, SCF-Skp2 and DDB1-Cul4, target human Cdt1 for proteolysis. *EMBO J.* 25, 1126–1136.
- Norio, P., Kosiyatrakul, S., Yang, Q., Guan, Z., Brown, N.M., Thomas, S., Riblet, R., and Schildkraut, C.L. (2005). Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol. Cell* 20, 575–587.
- Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl. Acad. Sci. U.S.A.* 59, 598–605.
- Olavarrieta, L., Hernández, P., Krimer, D.B., and Schvartzman, J.B. (2002). DNA knotting caused by head-on collision of transcription and replication. *J. Mol. Biol.* 322, 1–6.
- Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escribano-Diaz, C., and Durocher, D. (2014). Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. *Science* 344, 189–193.

- Panier, S., and Boulton, S. (2014). Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol* 15, 7–18.
- Patel, P.K., Kommajosyula, N., Rosebrock, A., Bensimon, A., Leatherwood, J., Bechhoefer, J., and Rhind, N. (2008). The Hsk1(Cdc7) replication kinase regulates origin efficiency. *Molecular Biology of the Cell* 19, 5550–5558.
- Petersen, BO, Lukas, J, and Sørensen, CS (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *EMBO J.* 18(2):396-410.
- Picard, F., Cadoret, J.-C.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.-N.N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genetics* 10, e1004282.
- Pillaire, M.-J.J., Selves, J., Gordien, K., Gourraud, P.-A.A., Gouraud, P.-A.A., Gentil, C., Danjoux, M., Do, C., Negre, V., Bieth, A., et al. (2010). A “DNA replication” signature of progression and negative outcome in colorectal cancer. *Oncogene* 29, 876–887.
- Poli, J., Tsaponina, O., Crabbé, L., Keszthelyi, A., Pantesco, V., Chabes, A., Lengronne, A., and Pasero, P. (2012). dNTP pools determine fork progression and origin usage under replication stress. *EMBO J.* 31, 883–894.
- Prasanth, S.G., Méndez, J., Prasanth, K.V., and Stillman, B. (2004). Dynamics of pre-replication complex proteins during the cell division cycle. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 359, 7–16.
- Pursell, Z.F., Isoz, I., Lundström, E.-B.B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* 317, 127–130.
- Radichev, I., Parashkevova, A., and Anachkova, B. (2005). Initiation of DNA replication at a nuclear matrix-attached chromatin fraction. *Journal of Cellular Physiology* 203(1):71-7
- Remus, D., Beall, E.L., and Botchan, M.R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J.* 23, 897–907.
- Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell* 139, 719–730.
- Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harbor Perspectives in Biology* 5, a010132.
- Roerink, S.F., van Schendel, R., and Tijsterman, M. (2014). Polymerase theta-mediated end joining of replication-associated DNA breaks in *C. elegans*. *Genome Res.* 24, 954–962.
- Ryba, Battaglia, Chang, Shirley, Buckley, Pope, Devidas, Druker, and Gilbert (2012). Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Research* 22, 1833-1844.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 20, 761–770.
- Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I., and Gilbert, D.M. (2011a). Genome-scale analysis of replication timing: from bench to bioinformatics. *Nat Protoc* 6, 870–895.
- Ryba, T., Hiratani, I., Sasaki, T., Battaglia, D., Kulik, M., Zhang, J., Dalton, S., and Gilbert, D.M. (2011b). Replication timing: a fingerprint for cell identity and pluripotency. *PLoS Comput. Biol.* 7, e1002225.

- Sadoni, N., Cardoso, M.C., Stelzer, E.H., Leonhardt, H., and Zink, D. (2004). Stable chromosomal units determine the spatial and temporal organization of DNA replication. *J. Cell. Sci.* 117, 5353–5365.
- Sasaki, T., Sawado, T., Yamaguchi, M., and Shinomiya, T. (1999). Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNAPolalpha-dE2F locus of *Drosophila melanogaster*. *Molecular and Cellular Biology* 19, 547–555.
- Seki, M., Marini, F., and Wood, R.D. (2003). POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. *Nucleic Acids Res.* 31, 6117–6126.
- Seki, M., Masutani, C., Yang, L.W., Schuffert, A., Iwai, S., Bahar, I., and Wood, R.D. (2004). High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J.* 23, 4484–4494.
- Senga, T., Sivaprasad, U., Zhu, W., Park, J., Arias, E.E., Walter, J.C., and Dutta, A. (2006). PCNA is a cofactor for Cdt1 degradation by CUL4/DDB1-mediated N-terminal ubiquitination. *Journal of Biological Chemistry* 281, 6246–6252.
- Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genetics* 5, e1000446.
- Sheehan, M.A., Mills, A.D., Sleeman, A.M., Laskey, R.A., and Blow, J.J. (1988). Steps in the assembly of replication-competent nuclei in a cell-free system from *Xenopus* eggs. *The Journal of Cell Biology* 106, 1–12.
- Shima, N., Munroe, R.J., and Schimenti, J.C. (2004). The mouse genomic instability mutation *chaos1* is an allele of *Polq* that exhibits genetic interaction with *Atm*. *Mol. Cell. Biol.* 24, 10381–10389.
- Shinbrot, E., Henninger, E.E., Weinhold, N., Covington, K.R., Göksenin, A.Y., Schultz, N., Chao, H., Doddapaneni, H., Muzny, D.M., Gibbs, R.A., et al. (2014). Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.* 24, 1740–1750.
- Siddiqui, K., On, K.F., and Diffley, J.F. (2013). Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* 5.
- Smith, L., Plug, A., and Thayer, M. (2001). Delayed replication timing leads to delayed mitotic chromosome condensation and chromosomal instability of chromosome translocations. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13300–13305.
- Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* 25, 514–522.
- Sollier, J., Stork, CT, García-Rubio, ML, and Paulsen, RD (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell.* 56(6):777-85
- Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat. Struct. Mol. Biol.* 12, 965–971.
- Stamatoyannopoulos, JA, Adzhubei, I, and Thurman, RE (2009). Human mutation rate associated with DNA replication timing. *Nat Genet.* 41(4):393-5.
- Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* 391, 231–232.
- Stillman, B. (2015). Reconsidering DNA Polymerases at the Replication Fork in Eukaryotes. *Mol. Cell* 59, 139–141.
- Sudakin, V., Ganioth, D., Dahan, A., Heller, H., Hershko, J., Luca, F.C., Ruderman, J.V., and Hershko, A. (1995). The cyclosome, a large complex containing cyclin-selective ubiquitin ligase activity, targets cyclins for destruction at the end of mitosis. *Mol. Biol. Cell* 6, 185–197.

- Sun, J., Kawakami, H., Zech, J., Speck, C., Stillman, B., and Li, H. (2012). Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure* 20, 534–544.
- Le Tallec, B., Dutrillaux, B., Lachages, A.-M.M., Millot, G.A., Brison, O., and Debatisse, M. (2012). Molecular profiling of common fragile sites in human fibroblasts. *Nature Structural & Molecular Biology* 18, 1421–1423.
- Tanaka, S., and Diffley, J.F. (2002). Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. *Nat. Cell Biol.* 4, 198–207.
- Tanaka, S., Nakato, R., Katou, Y., Shirahige, K., and Araki, H. (2012). Origin association of Sld3, Sld7, and Cdc45 proteins is a key step for determination of origin-firing timing. *Current Biology : CB* 21, 2055–2063.
- Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. *Nat. Cell Biol.* 12, 1086–1093.
- Tazumi, A., Fukuura, M., Nakato, R., Kishimoto, A., Takenaka, T., Ogawa, S., Song, J.-H.H., Takahashi, T.S., Nakagawa, T., Shirahige, K., et al. (2012). Telomere-binding protein Taz1 controls global replication timing through its localization near late replication origins in fission yeast. *Genes Dev.* 26, 2050–2062.
- Ticau, S, Friedman, LJ, Ivica, NA, Gelles, J, and Bell, SP (2015). Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. *Cell.* 161(3):513-25
- Toledo, L.I., Altmeyer, M., Rask, M.-B.B., Lukas, C., Larsen, D.H., Povlsen, L.K., Bekker-Jensen, S., Mailand, N., Bartek, J., and Lukas, J. (2013). ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* 155, 1088–1103.
- Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* 11, 1315–1324.
- Valton, A.-L.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintomé, C., Riou, J.-F.F., and Prioleau, M.-N.N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 33, 732–746.
- Van, C., Yan, S., Michael, W.M., Waga, S., and Cimprich, K.A. (2010). Continued primer synthesis at stalled replication forks contributes to checkpoint activation. *J. Cell Biol.* 189, 233–246.
- Vashee, S., Cvetič, C., Lu, W., Simancek, P., Kelly, T., and Walter, J. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes & Development* 17, 1894–1908.
- Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews. Molecular Cell Biology* 13, 115–126.
- Watanabe, Y., Fujiyama, A., Ichiba, Y., Hattori, M., Yada, T., Sakaki, Y., and Ikemura, T. (2002). Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Hum. Mol. Genet.* 11, 13–21.
- Wilhelm, T., Magdalou, I., Barascou, A., Técher, H., Debatisse, M., Lopez, BS. (2014) Spontaneous slow replication fork progression elicits mitosis alterations in homologous recombination-deficient mammalian cells. *PNAS* 111(2):763-8.
- Wilson, R.H., and Coverley, D. (2013). Relationship between DNA replication and the nuclear matrix. *Genes to Cells : Devoted to Molecular & Cellular Mechanisms* 18, 17–31.

- Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science* 290, 2309–2312.
- Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I., and Carter, N.P. (2004). Replication timing of the human genome. *Human Molecular Genetics* 13, 191–202.
- Wu, J.R., and Gilbert, D.M. (1996). A distinct G1 step required to specify the Chinese hamster DHFR replication origin. *Science (New York, N.Y.)* 271, 1270–1272.
- Wu, P.-Y.J.Y., and Nurse, P. (2009). Establishing the program of origin firing during S phase in fission Yeast. *Cell* 136, 852–864.
- Xiong, X., Du, Z., Wang, Y., Feng, Z., Fan, P., Yan, C., Willers, H., and Zhang, J. (2015). 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. *Nucleic Acids Research* 43, 1659–1670.
- Yamazaki, S., Ishii, A., Kanoh, Y., Oda, M., Nishito, Y., and Masai, H. (2012). Rif1 regulates the replication timing domains on the human genome. *EMBO J.* 31, 3667–3677.
- Yang, S.C., Rhind, N., and Bechhoefer, J. (2010). Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. *Molecular Systems Biology* 6, 404.
- Yeeles, J.T., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature* 519, 431–435.
- Yekezare, M., Gómez-González, B., and Diffley, J.F. (2013). Controlling DNA replication origins in response to DNA damage - inhibit globally, activate locally. *Journal of Cell Science* 126, 1297–1306.
- Ying, S., Minocherhomji, S., Chan, K.L., Palmai-Pallag, T., Chu, W.K., Wass, T., Mankouri, H.W., Liu, Y., and Hickson, I.D. (2013). MUS81 promotes common fragile site expression. *Nat. Cell Biol.* 15, 1001–1007.
- Yoon, J.-H.H., Roy Choudhury, J., Park, J., Prakash, S., and Prakash, L. (2014). A role for DNA polymerase θ in promoting replication through oxidative DNA lesion, thymine glycol, in human cells. *J. Biol. Chem.* 289, 13177–13185.
- Yoshimura, M., Kohzaki, M., Nakamura, J., Asagoshi, K., Sonoda, E., Hou, E., Prasad, R., Wilson, S.H., Tano, K., Yasui, A., et al. (2006). Vertebrate POLQ and POLbeta cooperate in base excision repair of oxidative DNA damage. *Mol. Cell* 24, 115–125.
- Yousefzadeh, M.J., and Wood, R.D. (2013). DNA polymerase POLQ and cellular defense against DNA damage. *DNA Repair (Amst.)* 12, 1–9.
- Yousefzadeh, M.J., Wyatt, D.W., Takata, K.-I., Mu, Y., Hensley, S.C., Tomida, J., Bylund, G.O.O., Doublié, S., Johansson, E., Ramsden, D.A., et al. (2014). Mechanism of suppression of chromosomal instability by DNA polymerase POLQ. *PLoS Genet.* 10, e1004654.
- Yüce, Ö., and West, S.C. (2013). Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol. Cell. Biol.* 33, 406–417.
- Zahn, K.E., Averill, A.M., Aller, P., Wood, R.D., and Doublié, S. (2015). Human DNA polymerase θ grasps the primer terminus to mediate DNA repair. *Nat. Struct. Mol. Biol.* 22, 304–311.
- Zan, H., Shima, N., Xu, Z., Al-Qahtani, A., Evinger Iii, A.J., Zhong, Y., Schimenti, J.C., and Casali, P. (2005). The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. *EMBO J.* 24, 3757–3769.
- Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. *Nat. Cell Biol.* 16, 2–9.

- Zhou, B.B., and Elledge, S.J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature* 408, 433–439.
- Zimmermann, M., and de Lange, T. (2013). 53BP1: pro choice in DNA repair. *Trends Cell Biol.* 24(2):108-17.
- Zimmermann, M, Lottersberger, F, Buonomo, SB, and Sfeir, A (2013). 53BP1 regulates DSB repair using Rif1 to control 5' end resection. *Science*. 339(6120):700-4
- Zlotorynski, E, Rahat, A, and Skaug, J (2003). Molecular basis for expression of common and rare fragile sites. *Mol Cell Biol.* 23(20):7143-51.

RÉSUMÉ

La duplication de l'ADN au cours de la phase S est initiée à partir de l'activation de plusieurs dizaines de milliers d'origines de réplication. La mise en place des origines a lieu au cours de la phase G1 sous la forme de complexe de pré-réplication (pré-RC) et leur activation est orchestrée par un programme spatio-temporel. La régulation spatiale détermine les origines qui seront activées et la régulation temporelle, ou *timing* de réplication, détermine le moment de leur activation. En effet, toutes ces origines ne sont pas activées en même temps durant la phase S : certaines origines seront activées en début de phase S, d'autre en milieu, ou d'autre à la fin. Ce programme est établi en tout début de phase G1, au « point de décision du *timing* ». C'est un programme très robuste qui signe l'identité d'une cellule, son état de différenciation et le type cellulaire à laquelle elle appartient. Il a aussi été montré qu'il est altéré dans des situations pathologiques, en particulier le cancer, sans qu'on ne comprenne très bien les raisons mécanistiques. De manière générale, les mécanismes moléculaires qui régulent le *timing* de réplication sont méconnus.

Le premier volet de ma thèse a permis l'identification d'un nouveau régulateur du *timing* de réplication : il s'agit de l'ADN polymérase spécialisée Thêta. Recrutée à la chromatine très tôt en phase G1, elle interagit avec des composants du pré-RC, et régule le recrutement des hélicases réplcatives à la chromatine. Enfin, sa déplétion ou sa surexpression entraîne une modification du *timing* de réplication à l'échelle du génome.

Dans la deuxième partie de ma thèse, j'ai exploré les mécanismes qui régulent ce programme temporel d'activation des origines suite à un stress réplcatif. J'ai identifié un mécanisme de régulation transgénérationnel inédit : la modification du *timing* de réplication de domaines chromosomiques ayant subi un stress réplcatif au cycle cellulaire précédent. Des cellules-filles issues d'une cellule ayant subi des problèmes de réplication dans des domaines fragiles (riches en AT, et donc potentiellement structurés, et pauvres en origines) présentent un *timing* plus précoce de l'activation des origines au niveau de ces domaines. Ce nouveau processus biologique d'adaptation est particulièrement intéressant dans un contexte tumoral de haut stress réplcatif chronique car ce pourrait être un moyen pour la cellule tumorale de survivre à son propre stress réplcatif mais aussi aux thérapies antitumorales qui sont nombreuses à cibler la réplication de l'ADN.

ABSTRACT

DNA duplication in S phase starts from thousands of initiation sites called DNA replication origins. These replication origins are set in G1 as pre-replication complexes (pre-RC) and fired in S phase following a spatio-temporal program of activation. This program determines which origins will be fired and when. Indeed, all the origins are not fired in the same time and we can distinguish early, middle and late replication origins. This temporal regulation is called "replication timing" and is determined at the "timing decision point" (TDP) in early G1. It's a robust program, which participates to the definition of cell identity, in term of differentiation state or cell type. However, the precise molecular mechanisms involved are poorly understood. Defective timing program has been evidenced in pathological contexts, in particular in cancers, but the mechanisms of this deregulation remain unclear.

In the first part of my PhD, I contributed to the discovery of a new regulator of the origin timing program: the specialized DNA polymerase Theta (Pol θ). Pol θ is loaded onto chromatin in early G1, coimmunoprecipitates with pre-RC components and modulates the recruitment of Mcm helicases at TDP. Moreover, depletion or overexpression of Pol θ modifies the timing of replication at a fraction of chromosomal domains.

The second part of my work aimed at exploring the mechanisms that regulates replication timing after a replicative stress. I identified a totally new transgenerational adaptive mechanism of DNA replication timing regulation: the modification of the timing of origin activation at chromosomal domains that have suffered from a replicative stress during the previous cell cycle. Daughter cells from a cell that has experienced replication stress at particular domains (late replicating domains, AT rich so they can form structured DNA, and poor in origin density) shows advanced origin activation within these regions. This new biological process in response to replicative stress could be of particular interest in the context of cancer since, tumor cells are characterized by high level of intrinsic chronic replicative stress. This new mechanism may favor cancer cell survival despite replication stress, particularly upon treatments with anti-tumor agents that target DNA.